NATURE VOL 341 nature 19 OCTOBER 1989

Wall Street's second tumble

Coincidences are usually meaningless, but can it be significant that last Friday's tumble on Wall Street marks almost exactly the second anniversary of the crash on 19 October 1987, a Monday? Magic is no part of the explanation.

ON the assumption that the failure of two US banks to raise the loans required to pay part of the purchase price for United Airlines was the occasion for, but not the cause of, last Friday's sudden collapse on Wall Street, the belief that October is a dangerous time of the year is bound to gain currency. But what can be the explanation? That October marks the beginning of the US government's fiscal year may not be entirely irrelevant, for it is about now each year that the United States usually discovers that the administration and the Congress have reached an unsatisfactory compromise on spending for the year ahead. This year, of course, there is no compromise, but President Bush will have to announce that all discretionary elements of federal spending (including research support) will have to be reduced by roughly 1.5 per cent. This decision, also reached last Friday, can hardly have inspired confidence in Wall Street, or anywhere else.

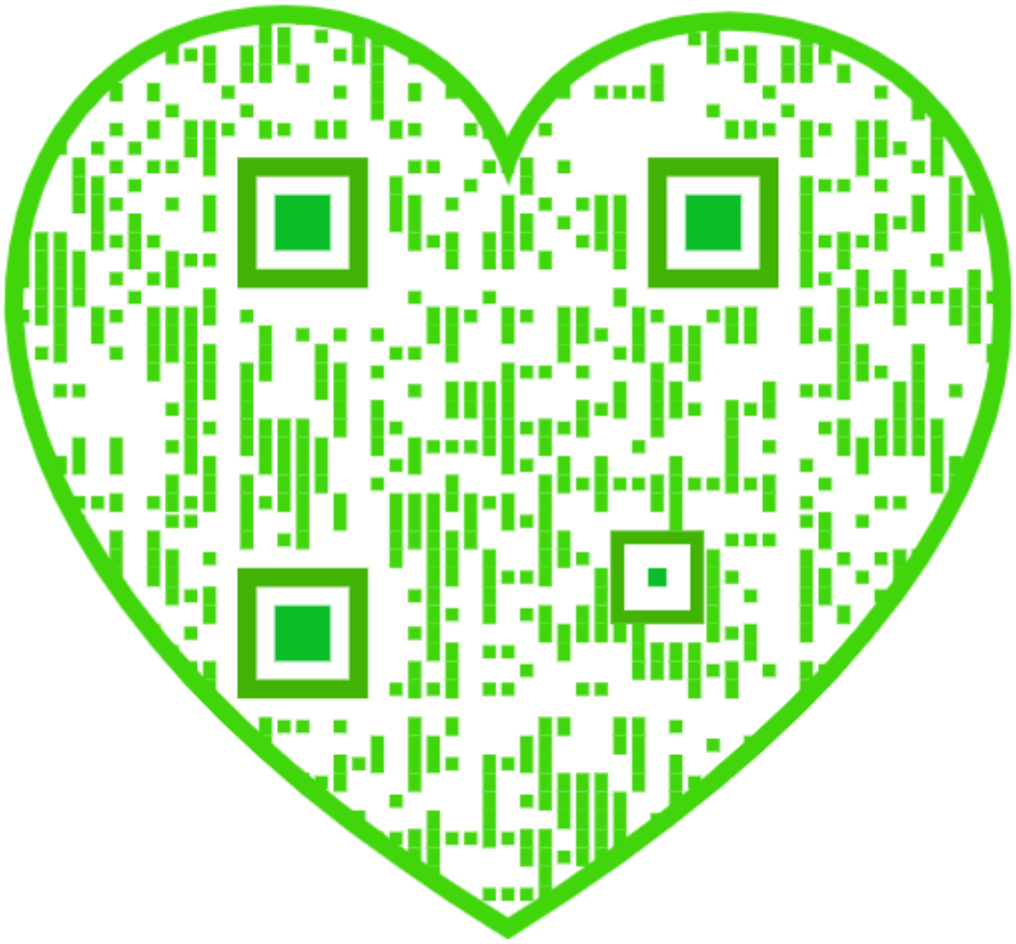
What else of significance can there be? In 1987, as this year, the International Monetary Fund had just met in Washington, uttering a largely cheerful prognosis for the years ahead. Then, as this year, the fragility of the world's financial system engendered by the $300,000 millions of debts owed by developing countries was underplayed, which is hardly a confidence-building measure. In a separate meeting, the finance ministers and central bankers of the seven important industrialized countries, in 1987 as this year, offered a comforting account of the progress of the world's economy. If the bankers, who should know best, fail to recognize the fragility of the markets, how can the traders who live in them go about their business confidently? If these financial institutions cannot change their ways, they might at least change the season of their meetings, say to the northern spring, so that it will be possible for the rest of us to judge empirically the extent to which their persisting blandness scares the money people out of their wits.

But it is pointless to search for an explanation of coincidence when the causes of last week's tumble are persistent and plain. Since the beginning of this year, there has been a steady trickle of depressing news about the market in junk bonds — the financial instruments which have flourished in the past few years and which, unlike other bonds, are secured not on tangible assets but on the future income of a company. By means of junk bonds, almost anybody can set out to buy virtually any publicly quoted company; it is simply necessary to sell enough debt to cover the purchase price. In practice, the owners of a company will usually not sell unless they are offered some cash along with the junk.

If things go well, the purchaser can hope to make a profit by selling off pieces of his purchase, while the holders of the junk bonds enjoy the high interest appropriate to risky investments. Sadly, for some of them, if things do not go well, everybody catches cold. The largest casualty this year is the Canadian company called Campeau, which bought a chain of US retail stores with $10,000 million of junk bonds and then found, two months ago, that it could not pay the interest to which it is commited and also fill its shops with goods for the Christmas spending season. The spectre Wall Street's traders glimpsed last Friday was that of the decimation of this huge amount of debt, comparable in amount with the $300,000 million the poor countries of the world owe their creditor banks. The trouble about United Airlines was that banks were unwilling to provide the cash that would earlier have been provided by the junk bond market if it had not dried up.

So what, it may be tempting to ask; what does it matter if financial speculators lose their shirts from time to time? One answer is that it is not their money, but, in some sense or another, ours. The financial institutions that keep the stock markets alive are mostly the repositories, in the industrialized countries of the West, of ordinary people's savings. Second, the steps that governments will now be compelled to take to remedy the effects of the crash that began on Friday are bound to hurt. In the United States, for example, it seems increasingly unlikely that President Bush will be able to hold to his declaration that taxes will not be increased. The central bank (the Federal Reserve) will be compelled to reduce interest rates to combat the deflationary effect of the tumble, increasing the difficulty of tempting people to cover the federal deficit with their savings. In Britain, for much the same reasons, it is less and less likely that the government will be able to continue pretending that the pound sterling is worth almost exactly 3.0 deutschmarks, and that may be no bad thing, but one consequence is that inflation will be increased.

These events will also exacerbate the long-standing distrust of the financial communities by industry, forever complaining that people are more ready to invest in funny financial instruments than in productive assets. Often, the



complaint is self-serving, the defence of inefficient management. But what this second tumble on Wall Street shows is that the short-term calculations of the financial communities that there is more profit in the junk bond market and other expedients for making money than in industrial investment are somehow mistaken. The West being what it is, the only long-term remedy must be to provide incentives for productive industrial investment. Then, there would be more financial securities to soak up people's savings, so that stock market prices would settle at more realistically low levels, while interest rates would decline with people's expectations of what constitutes a sustainable return on their investments. But that halcyon dream is separated from the present by the huge uncertainties of the weeks that lie ahead. □

Agricultural contest

The US Department of Agriculture should respond to a plea to fund research competitively, but with guille.

ALTHOUGH it is high time that an influential body such as the National Research Council (NRC) complained that the US Department of Agriculture (USDA) should spend its research funds more intelligently (see page 561), it should not be surprised if its arguments fail to be reflected in the shape of the federal budget. There have been some occasions in the past forty years when the department itself has been persuaded that it should change its ways, these have never coincided decisively with occasions when the Congress is prepared to sanction change. The stumbling-block is always that the Congress, but sometimes also USDA itself, seems to be persuaded that spending more on competitive research grants would be the death of the traditional way of of channelling funds to US universities — the hallowed but often abused system of land-grant colleges.

It need not be like that at all. The land-grant system, originally a scheme for stiffening the agricultural extension services intellectually, is rightly respected throughout the world. The original calculation was that the generous endowment of agriculture and related departments at universities, and the creation of links with federal and state extension services, would give the United States a productive and resilient agricultural economy. No disrespect of USDA's own research establishments is implied by the acknowledgement that the original calculations were accurate. US agriculture is indeed the most productive in the world.

That is the good news. The other side of the coin is that times have changed and that, among the agricultural establishments of the land-grant colleges, some have turned out to be better than others. While nobody can fault, for example, the University of California at Davis for its eagerness to seize on new techniques such as those suggested by molecular biology, others are still rather woodenly concerned with the applied research that used to be their staple diet, but which is increasingly irrelevant to the needs of US agriculture. It is especially galling that many of the sleepy but rich places live alongside people in other than agricultural departments who have to fight their way into the competitive world of modern biology through the eyes of all the needles that make up the competitive grants business administered by the NIH.

What NRC asks is that there should be an extra $500 million — more than ten times present spending — for a new programme of competitive grants in agriculture. Its calculation that the investment would pay off is probably correct. The practical and political problem is whether the new programme could be introduced in such a way that the old agricultural establishments would not be left entirely out in the cold; otherwise, there is an eager band of congressmen ready to leap to their defence. Tactically, this requires some kind of understanding between university presidents and USDA on how to revivify the places now asleep. □

Paying academics

British university teachers are ill-paid, but their new salary claim should be more subtle.

THE British university teachers' union, the Association of University Teachers (AUT), resolved at the week-end to ask for a 27 per cent salary increase. In equity, the claim is undeniable. Indeed, the erosion of university teachers' salaries by comparison with those their graduates quickly earn has been so rapid that it is surprising that the claim is so relatively modest. One difficulty is that their employers cannot afford to meet the claim; any extra funds will have to be provided by the government as part of its general subvention of the universities. Another is that university teachers will undermine a good case if they again engage, as threatened, in a boycott of university examinations or some other such unprofessional protest. A third, is that the present system of national bargaining on university salaries is an anachronism

There are good tactical reasons why AUT as such cannot confess that the time has come when teachers at one university should not expect to be paid what those at another earn; to concede that at the outset of what must be a bruising negotiation (from which the real paymaster will be absent) would evidently undermine AUT's case. But entirely to deny the likelihood that there will eventually be a time when universities compete for teachers, and when the stronger get the best, is unrealistic.

Paradoxically, it could be in the best interests of most teachers at British universities that national agreements should be abandoned. Already, there is resentment in the universities that teachers at British polytechnics are better paid. But just as national salary scales are in the long run untenable, so is the division between universities and polytechnics. Universities (and academics) being what they are, it would not be in the least surprising that, when the system settles down, Oxbridge salaries will be less than at the University of the Stix. □

Is the end of particle proliferation at hand?

- Only three particle families exist
- Rubbia accuses SLAC of sharp practice

London & Washington

IN parallel announcements from the Stanford Linear Accelerator Center (SLAC) in California and the European Particle Physics Laboratory, CERN, in Geneva, physicists proclaimed last week that they had put a limit on the number of 'families' of elementary particles.

The two laboratories, which house the world's principal positron–electron colliders, had been engaged in an informal race to establish the new result, but CERN's director, Carlo Rubbia, has accused SLAC of using "hockey player's tactics" and of "scraping the bottom of the barrel" in making an announcement that beat CERN's by one day.

But Burton Richter, director of SLAC, says there was no contrivance in the near simultaneous declarations, and denies that his team went to any unusual lengths to "dig things out of the data" prematurely.

Both the Stanford Linear Collider (SLC) at SLAC and the Large Electron–Positron Collider (LEP) at CERN were designed as Z^0 "factories" in which electron–positron collisions would create quantities of the Z^0 particle, a crucial ingredient of particle physics for the discovery of which Rubbia won the Nobel prize in 1984.

SLC, a technically complex adaptation of an existing machine, has had some difficulties reaching its design performance (*Nature* **340**, 587; 1989). LEP, by contrast, is a traditional circular synchrotron, and has come up to design specification only weeks after being switched on.

In the race to get big physics results, SLC got an early start, but LEP, by virtue of its higher beam energy, has produced more than 11,000 Z^0 particles in just 15 days, compared to 500 in 6 months of experiments at SLC. This mass-production has an important purpose. The known elementary particles, quarks and leptons, fall into three families or generations: each family includes two quarks, a heavy lepton, (the electron, muon or tauon) and the corresponding neutrino. The third family awaits the discovery of the top quark, but otherwise the picture is complete. But there has been some doubt whether there might not be one or more undiscovered families, and the lack of a clear answer has impeded theoretical understanding.

From the rate of production of Z^0s over a range of collision energies, it is possible to deduce the number of families: the greater the number, the more ways there are to make Z^0s and the greater is the "width" of the Z^0 — its production rate as a function of energy.

The new data from both CERN and SLAC show conclusively that three families is all there are, but this harmonious conclusion has been marred by Rubbia's allegations about last Thursday's press announcement from SLAC. CERN announced an identical conclusion, based on a greater volume of data, the next day.

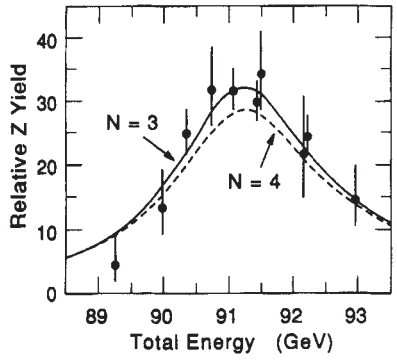
Speaking at a conference in Pisa at the weekend, Rubbia was evidently irritated at the order of the announcements, and sought to put the US group in what he regarded as its proper place. According to a report in the Italian newspaper *La Republica*, he said that "at best the Americans will confirm our data, coming in second place — a position they will never accept". The complaint is that SLC physicists knew of CERN's intention to announce the new LEP results last Friday and decided to scoop the Europeans by issuing the Stanford data a day earlier.

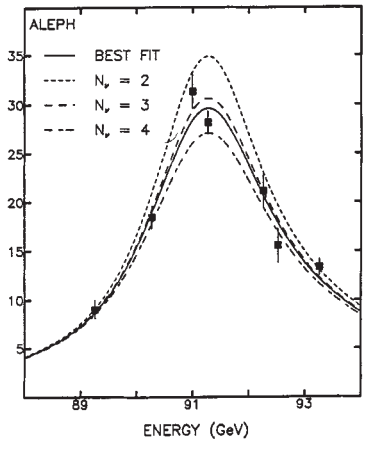
Richter, although hesitant to respond to newspaper reports of what Rubbia had said, denied there had been any attempt to steal CERN's thunder.

SLAC, he said, has been releasing progress reports on its work, and at the beginning of August had enough data to put a preliminary limit of 3.0 ± 0.9 on the number of families, which excludes the existence of four families with a confidence level of about 70 per cent. By the beginning of September, a better limit of 2.7 ± 0.7 had been obtained, excluding four with a confidence level of 96 per cent. This was announced at a meeting of the European Physical Society in Madrid. Then two weeks ago, a paper with this conclusion was submitted to *Physical Review Letters*, a press announcement was made and copies of the paper were faxed to, among other places, CERN.

Not everyone at CERN shares Rubbia's anger. Indeed, many feel that the scramble for press publicity, which seems to be a consequence of the struggle to find money for new accelerators, in the end does a disservice to high-energy physics by provoking public quarrels. Even Rubbia confessed that the race to find the elementary structure of matter is becoming "a soap opera — like Dallas".

Roland Pease & David Lindley





SLAC's version (above) and CERN's, the latter due to appear in *Phys. Rev. Lett.* B Vol. 231 (4), dated 16 November.

IVORY TRADE

African dissent at two-year ban

Paris

AFTER a week of debates, delegates at the seventh conference at Lausanne of the parties of the Convention on International Trade in Endangered Species of Wild Flora and Fauna (CITES) voted late on Monday to move the African elephant to Appendix 1 of the convention for two years. This in effect, means a blanket ban in trade in ivory from the African elephant. But nine African nations voted against the proposal, and five are to apply to enter a reservation which could mean that the ban will not apply to them.

The CITES decision has not pleased some African nations, such as Zimbabwe, who say that ivory trade is necessary for their economy and to control growing populations of elephant. Zimbabwe has developed a model practice for management of elephant culls and says that a ban will make it difficult to continue to make the policy work. But the CITES ban allows African nations to reapply for a licence to export ivory if, after two years, they can show that they have developed an effective strategy to prevent further decline in their elephant populations.

Over the past nine years, the population of elephants in Africa has declined from an estimated 1.3 million to around 600,000. According to some estimates more than 90 per cent of ivory leaving Africa is from animals killed by poachers. Peter Coles

NOBEL PRIZES: CHEMISTRY

RNA's catalytic role

London

THIS year's Nobel prize for chemistry has been awarded jointly to Sidney Altman of Yale University and Thomas Cech of the University of Colorado, Boulder, for their independent discoveries that RNA can have catalytic activity. These discoveries may have profound implications for our understanding of the origin of life, and may also lead to important practical developments.

Biological catalysis had previously been considered the exclusive domain of protein enzymes. RNA is usually thought of as a passive informational molecule, though specific RNA molecules have long been known to have non-informational roles in protein synthesis.

One of the central problems in understanding how life could have originated is posed by the interdependence of modern systems of informational molecules, such as DNA, and functional molecules such as proteins. Consideration of this problem led to speculation by the pioneers of molecular biology in the 1960s, most notably by Francis Crick, that catalytic RNA combining informational and functional properties might provide a solution to the seemingly intractable 'chicken-and-egg' problem. But such ideas were pure speculation until Altman and Cech, both working on RNA processing systems, demonstrated that RNA can indeed have catalytic properties similar to those of protein enzymes.

Altman's work concerned the bacterial enzyme ribonuclease P, which catalyses a specific cleavage reaction in the synthesis of mature tRNA molecules from larger precursors. In the 1970s Altman and colleagues had shown that ribonuclease P has RNA and protein components, both of which are essential for its enzymatic activity. They thought the RNA may be involved in recognizing the RNA substrates of ribonuclease P, but at first did not consider a catalytic role for the RNA component to be a serious possibility.

Meanwhile, Thomas Cech and his colleagues were investigating the process by which the 'intervening sequence' of the precursor of a ribosomal RNA of the protozoan *Tetrahymena thermophila* is excised and the flanking 'exons' joined to make the mature ribosomal RNA product.

In 1982 Cech's group obtained conclusive evidence that this excision and ligation process can occur spontaneously in the absence of any protein catalyst (*Cell* **31**, 147–157; 1982). Although not true catalysis, Cech's discovery provided the first firm evidence that RNA may be capable of specific catalysis. Altman and colleagues went on to find conditions in which the RNA component of ribonuclease P could, in the absence of protein, catalyse the maturation of tRNA (*Cell* **35**, 849–857; 1983).

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Celebrations Colorado style, Thomas Cech (left) and at Yale, Sidney Altman. (AP.)

Following this pioneering work many examples of RNA catalysis have been discovered. In 1986 Cech's group showed that part of the RNA excised from the *Tetrahymena* ribosomal RNA precursor has catalytic activity comparable to that of protein enzymes. Among the variety of reactions catalysed by this RNA — all variations on the basic theme of transesterification — one, synthesis of polycytidylic acid from pentacytidylic acid substrate, bore a striking resemblance to RNA polymerization.

A further step towards realizing Crick's speculation of an RNA molecule with replicase properties has recently been taken by the construction of a derivative of Cech's *Tetrahymena* RNA catalyst that catalyses the ligation of multiple oligonucleotides aligned on an external template (*Nature* **339**, 519–522; 1989).

The practical consequences inherent in these discoveries stem from the prospect of being able to manipulate specific RNA molecules in cells with RNA catalysts. An important step towards this goal was achieved last year by the construction of RNA catalysts derived from 'self-cleaving' plant-virus satellite RNA capable of cleaving other RNA molecules at highly specific, predetermined positions. This opens the possibility of using RNA catalysts to block the expression of specific genes in cells and organisms, may also lead to a new way of preventing virus infections, and has already led to a patents dispute involving Cech (*Nature* **341**, 473; 1989).

Geoffrey North

NOBEL DISPUTE

French researcher asks for share

Paris

LAST week's award of the Nobel prize in medicine to the Americans, J. Michael Bishop and Harold E. Varmus (see *Nature* 341, 475; 12 October 1989), has been greeted with sourness in France.

Dominique Stehelin, who is now director of research at a National Scientific Research Centre (CNRS) laboratory at the Pasteur Institute in Lille, said last week that it was he who carried out the work that led to the demonstration of the cellular character of oncogenes. A postdoctoral fellow in Bishop and Varmus' laboratory in the early 1970s, Stehelin was the first author of papers in *Nature* and the *Journal of Molecular Biology* announcing the discovery.

"I did all the work from A to Z", Stehelin told Agence France Presse on hearing that he would not share the Nobel with Bishop and Varmus.

Stehelin has since softened his position, saying that he does not contest the decision of the Nobel jury, but still wonders why they could not have added a third name to the prize. Meanwhile, the director-general of CNRS, Francois Kourilsky, issued a statement congratulating Bishop and Varmus on the award but adding that "on this occasion, CNRS wishes to salute the role played by one of its researchers, Dominique Stehelin, in this discovery".

This role was sufficiently significant, says the statement, to earn Stehelin the Rosen and Griffuel prizes and the CNRS silver medal. Peter Coles

■ Bishop and Varmus have not commented directly on Stehelin's complaint, but numerous colleagues have come to their defence. According to the *Washington Post*, Peter Vogt, the fourth author on the *Nature* paper, said that Stehelin's unhappiness was "understandable but regrettable". A statement issued by the University of California at San Francisco made it clear that Stehelin's work was done "under supervision". □

NOBEL PRIZES: PHYSICS

Measures towards success

Washington

THIS year's award of the Nobel prize for physics, half going to Norman F. Ramsey of Harvard University and half shared by Hans G. Dehmelt of the University of Washington in Seattle and Wolfgang Paul of the University of Bonn, has its roots in the beginnings of quantum theory.

The technical accomplishments of all three prizewinners rely on the phenomenon, first recognized by Einstein, of stimulated emission, wherein an atom in an excited electronic state can be induced to fall to a state of lower energy by electromagnetic radiation of the frequency corresponding to the energy difference.

The prizewinners' innovations have been instrumental in testing the quantum theory of atoms with exquisite precision, and have also led to precise standards of measurement of general utility. The atomic clock, which depends on Ramsey's inventions, is now used to define one second of time as 9,192,631,770 oscillations of between hyperfine levels of caesium.

Ramsey's great achievement was the refinement of I. I. Rabi's method for inducing atomic transitions by passing excited atoms through a magnetic field and an electromagnetic radiation field of fixed frequency. Ramsey added a second radiation field, setting up an interference pattern with the first in which photons of the precise frequency needed to induce an atomic transition existed only in certain places. This restriction of the volume in which transitions could occur increased the coherence of the photons thus emitted.

A later development, by Ramsey and Daniel Kleppner, was the invention of the hydrogen maser, in which excited atoms are held in a suitably tuned cavity. The maser not only allowed direct high-precision studies of the structure of the hydrogen atom, but also became widely used as a reference source in, for example, interferometric techniques in radio-astronomy. The accurate frequency standard provided by the hydrogen maser allows observed radio signals at physically separated radiotelescopes to be compared, and from the differences high-resolution radio maps are calculated.

In Ramsey's work and its offshoots, precision is obtained by persuading many identical atoms to emit radiation coherently. In the work of Paul and Dehmelt, the goal has been to trap single ions or electrons in a suitable arrangement of electromagnetic fields, so that their properties may be studied directly. Paul's work began in the 1950s in Bonn, where he first invented methods to focus ions using multipolar magnetic fields (as is done in particle accelerators today), separate ions of different mass, and store them in traps of electric fields with a pervading radiofrequency radiation field.

The trap works by ingenious application of Einstein's rules for the emission and absorption of radiation. An ion moving towards a source of electromagnetic radiation can be excited to a higher level by picking up a photon of suitable frequency. But when it de-excites, it emits a photon whose frequency, with respect to the stationary apparatus, is slightly increased by the amount of the Doppler effect. The ion therefore emits a photon of higher frequency than that absorbed, and the extra energy is taken out of the ion's motion: in other words, it slows down. This principle works to cool ions, and to prevent them from moving out of the trapping volume.

Dehmelt and his colleagues developed a slightly different method, the Penning trap, which they used in 1973 to capture and hold a single electron. This has enabled Dehmelt's group to measure the electron's *g*-factor — the departure of its magnetic moment from the classical value — to an accuracy of a few parts in 1,000 million. Like Ramsey's work with the hydrogen maser, such measurements have tested (and so far confirmed) the correctness of the quantum theory of electrodynamics. The award of the Nobel prize to Ramsey, Paul and Dehmelt recognizes that physics, at bottom, is the art of measuring nature.

David Lindley

Physicists rewarded: Wolfgang Paul (left), Hans Dehmelt and Norman Ramsey. (AP.)

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

SCIENCE JOURNALS

Breaking the paper barrier

New Delhi

A 30-YEAR-OLD policy banning the publication of Indian editions of foreign science and technology periodicals in order to protect local publishers may soon be abandoned. The Indian Department of Science and Technology (DST) has been asked by the government to re-examine the issue, and has sent out a two-page questionnaire to a cross-section of the scientific community in India.

The present policy was instituted in the mid-1950s because the government feared that local editions brought out by foreign publishers would suffocate domestic journals. This was part of a general Indian policy to achieve self-reliance in science and technology that included import limitations and customs barriers to protect local technological products.

But five years ago the government made an about-turn, liberalizing the technology importation policy because it found that protectionism had led only to industrial stagnation. The government is now wondering if the publishing policy should be relaxed, for similar reasons: Indian publishing has not prospered, and the cost of importing foreign journals is an increasing burden.

The opinion poll is being carried out by *Current Science*, whose editor, S. Ramaseshan, says there are two extreme views on the issue. Behind the restrictive policy lies the idea that publication of foreign science and technology journals in India could "undermine not only the efforts to build up our own journals, but even the formation of a cohesive and committed scientific community".

But the opposing view is that local editions of foreign journals are essential if India's scientific and technological activities are to be on a par with international standards. Competition, Rameseshan believes, would "improve the quality of our science and our journals".

Some 15,000 scientists are expected to participate in the opinion poll, and DST hopes to make its recommendations to the government by the beginning of next year. Respondents will be asked whether or not they want a revision of the policy, and to list up to ten foreign journals that they would like see published in local editions.

For the time being, the French magazine *La Recherche* is the only foreign scientific publication appearing in India in English, but the demand for foreign science journals in a country with over 1,200 research establishments and 125 universities is something that overseas publishers are not likely to ignore.

K.S. Jayaraman

WORLD PSYCHIATRIC ASSOCIATION

Soviet's Freudian slip

Athens

THE Soviet All-Union of Psychiatrists and Narcologists (AUSPN) is preparing a special session on Freudian theory, and has already begun publishing Freud's collected works, Soviet psychiatrists told a press conference at the World Psychiatric Conference this week. Speaking on "Psychiatry and *Perestroika*", the Soviet team pointed out that (as with so many changes in the Soviet Union today), the reappearance of Freud, and other innovators of psychological theory, was a return to the "Leninist" policies of the 1920s, after decades in which the whole concept of psychoanalysis and the unconscious was republished as "pseudo-science". Soviet psychology and psychiatry were, they stressed, originally founded on the German school of the early 1920s, and the AUSPN now wanted to resume its contacts with Germans and other West European psychiatric associations.

"*Perestroika*," said Dr Alexsandr Karpov, of the USSR Ministry of Health, "like any great social event, has its effect on mental health too".

What the AUSPN was seeking, of course, was reinstatement in the World Psychiatric Association (WPA), from which (under its previous name of the All-Union Society of Psychiatrists and Neuropathologists it resigned in 1983 to avoid expulsion over the issue of the political use of psychiatry. The key issue, therefore, was whether or not Soviet psychiatry had mended its ways, or could be seen to have done so while its former leaders, such as Dr Marat Vartanyan and Dr N. Zharikov still held high office in the AUSPN. (Both of these were present in Athens; neither was at the press conference.) The Soviet team stressed that things are changing in Soviet psychiatry, urging that it would be better to look at a more positive future than to the painful events of the past. But human rights campaigners were distributing a letter just received from Aleksandr Podrabinek, of the unofficial Independent Psychiatric Association in Moscow, listing a number of new admissions over the past few months, which, the letter said, were made on political grounds.

The Soviet team argued that *perestroika* is a slow process, that there are now stronger legal guarantees of patients' rights, the "psychiatric register" of former patients has been greatly reduced, and that earlier this year, a "verification team" of US psychiatrists and lawyers had been given wide facilities to visit Soviet psychiatric hospitals.

The Soviet team, was challenged by Dr Anatolic Koryagin — a former Soviet psychiatrist who was himself incarcerated for having revealed (in *The Lancet*) details of Soviet abuses — about the AUSPN's letter of withdrawal from the WPA in 1983. If they recognized that political abuse of psychiatry had taken place in the past, he asked, how did they evaluate the signatures of the current leaders of Soviet psychiatry on the letter of resignation, which accused the UK Royal College of Psychiatrists of conducting a "slanderous campaign against them".

The President of the Royal College added his own demand for a formal apology; particularly as the chief signatory to the 1983 letter, Dr Georgii Morozov was, in 1988, elected an honorary president of the WPA — elected, it was stressed, by democratic, secret ballot and not imposed from above. The Soviet panel were reluctant to apologize on Georgii Morozov's behalf, but promised to inform him that an apology had been asked for. They would go no further than admitting that "mistakes were made in the past". But to the majority of press, psychiatrists and campaigners at the meeting, this answer was clearly insufficient. **Vera Rich**

RESEARCH FUNDS

MIT in the red

Boston

IN what could be a portent of difficult financial times ahead, the Massachusetts Institute of Technology (MIT) has posted its first budget deficit in six years. A recent university report discloses a shortfall of more than $5 million for the 1989 fiscal year, which ended in June. The deficit is largely a result of falling income at the Lincoln Laboratory, an MIT affiliate which is supported almost entirely by the Department of Defense (DoD).

This year's deficit could be early evidence of coming financial difficulties because of the university's substantial dependence on military funding. Kenneth A. Smith, vice president for research at the university, has predicted "significant cutbacks in military funds" for research by the 1991 fiscal year.

The deficit comes despite a small decrease in expenses from fiscal year 1988 which the university attributes to layoffs of subcontractors at Lincoln Laboratory. But operating revenues decreased faster. The resulting shortfall of more than $11 million was reduced to $5 million by consideration of nearly $6 million in unrestricted gifts and other income.

More worrying for the university is that the deficit has arisen in spite of a 19.3 per cent increase in support from non-federal sources. Fiscal year 1989 marked the first time that industry and other private funds surpassed the National Science Foundation as the fourth largest sponsor of research. The top three remained the same: the Department of Energy, the Department of Health and Human Services, and the DoD.

Despite rising concern, there are few solid proposals for replacing lost income. But Kosta Tsipis, director of the Program in Science and Technology for International Security and long a critic of MIT's reliance on military research dollars, is to hold a "high-level" meeting of corporate, academic, government and military leaders later this year to explore the prospects for converting some of the military-sponsored work at the university to more constructive research. Seth Shulman

GALILEO

Legal moves fail

Washington

AFTER surviving an eleventh-hour legal attack, the launch of the space shuttle *Atlantis*, carrying the Jupiter probe Galileo, had to be postponed from 12 October because of a technical problem. But by Saturday all was well again, and the launch had been rescheduled for early this week — Tuesday, 17 October. The unexpected delay gave the anti-nuclear opponents of Galileo's flight a chance to appeal against the earlier legal ruling, but that too failed.

The delay, announced by the National Aeronautics and Space Administration (NASA) only an hour after it had been given legal permission to continue the countdown, was due to a malfunction in a computer controlling one of the main shuttle engines. Although the computer has a back-up, mission controllers decided to replace the offending part. This was successfully done on Saturday, allowing a new lift-off time at 12.57 p.m. on the following Tuesday.

A request for an injunction against the launch of Atlantis, on the grounds that the environmental hazards from accidental dispersion of the plutonium in Galileo's power generators had been improperly assessed (see *Nature* **341**, 374; 1989), was turned down by Judge Oliver Gasch in the US District Court for the District of Columbia late last Tuesday.

Gasch declared that the court was not willing "to substitute its judgement ... for that of the government agency", a pronouncement that NASA took as a vindication of its risk assessment analysis but which Andrew Kimbrell, who argued for the injunction on behalf of the Foundation on Economic Trends, said was an abdication of the court's responsibility.

With legal recourse having failed, protesters may now attempt to make good their promise to prevent the launch by infiltrating the Kennedy Space Center and occupying the launch area. NASA has tightened security at the site, but is not publicly anticipating trouble.

David Lindley

COMPUTER ABUSE

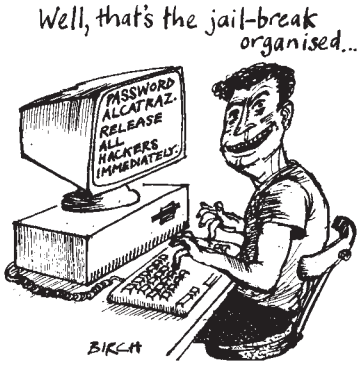
Plan to outlaw hacking

London

THE definition of three new crimes to deal with the growing problem of computer hacking, the unauthorized entry into computer systems, is recommended in a British Law Commission report released last week. The commission, a statutory body advising the British government on the legal aspects of new legislation, was asked to review the present legal status of hacking.

The commission says that British criminal law at present fails to cover two main problems: the "costs, disruption and uncertainty" caused to owners by unauthorized entry to their computer systems and the loss and damage caused by alterations of stored data and programs. "Such attacks can be particularly serious if they are directed against computers performing industrial or commercial operations", the report says.

The report, *Criminal Law — Computer Misuse*, concludes that an adequate deterrent cannot be constructed simply by



patching up existing criminal law, and recommends the definition of three new crimes. Any attempt to gain entry to a computer system, whether for gain or for fun, should be a criminal offence with a maximum penalty of three months in jail. Entry with intent to commit a crime should be treated more severely, with a maximum penalty of five years' imprisonment.

The commission recommends that there should be a third more serious offence consisting of attempting "intentionally and without authority" to alter data or programs with the use of worms, viruses or logic bombs, or of circulating disks infected with such devices. The maximum penalty for this offence would also be five years' imprisonment. The Trade and Industry Secretary, Mr Nicholas Ridley, welcomed the report and was "inclined to accept" its recommendations.

Law Commissioner Richard Buxton QC, who heads the Law Commission's criminal law team, said the system was designed to operate in a graded manner. The first offence would deter the "widespread activity" of hacking. This will be reinforced by the two more serious offences. Buxton added that the rules were carefully designed not to penalize computer users whose waywardness consists merely of carelessness. **Ben Webb**

■ The prosecution of Cornell University graduate student Robert T. Morris can succeed only by proving that he knowingly intended to gain unauthorized access to a number of federal computers with the virus that infected the Internet system a year ago (see *Nature* **340**, 329; 1989). It has been argued by some in the United States that any law that made unauthorized entry a crime, regardless of intent or authority, would be unworkable: in the Morris case, a number of network users put together "patches" that protected computers from infection and distributed them, in a freelance manner, across the network. Under a law that condemned Morris, would they also be guilty? □

MICROELECTRONICS

JESSI enters critical phase

Munich

As part of its commitment to the European microelectronics initiative known as JESSI (Joint European Submicron Silicon), West Germany announced on 28 September the foundation of a new institute for silicon technology. The institute will concentrate on the development of smaller memory chips, with a diameter of 0.5 micrometres or less, using X-ray lithography and other techniques.

JESSI, a collaboration of five European governments, the European Economic Commission and three of the largest electronics companies in Europe, is an ambitious attempt to combine research with product development across national borders (see *Nature* **339**, 652; 1989). The latest announcement reflects the strong West German commitment to JESSI at a time when critical decisions are being made concerning individual JESSI projects and possible US participation.

The Fraunhofer Society for the Advancement of Applied Science was expected to confirm its sponsorship of the institute at a meeting on 18 October. The new institute, which will be located in Itzehoe in the north German *Land* of Schleswig-Holstein, will differ from other Fraunhofer Institutes in that just a few large firms, including European chip producers such as Siemens and Philips, will participate in the research and benefit from the results.

The Research Ministry (BMFT) and the *Land* of Schleswig-Holstein will share the construction and equipment costs of DM 400 million ($210 million). When the institute begins operation, by 1992, industry will provide 20 per cent of the staff and will share in salaries and costs.

The Fraunhofer Institute for Microstructure in West Berlin, which has worked with X-ray lithography in the past, will lose some researchers to the Itzehoe institute. BMFT has announced that the Berlin institute will be restructured, causing some upset among West Berlin politicians, who had hoped that JESSI would help attract new high-technology investment to their isolated city.

The most important decision that is expected in the next few months concerns participation by the United States in JESSI. The Europeans have made the participation of firms such as IBM contingent on a reciprocal acceptance of European companies, or their US subsidiaries, into the US electronics consortium Sematech. Both JESSI and Sematech are founded on the principle that Japan is the main competition in designing and producing chips.

According to a spokesman, JESSI representatives paid a visit on Sematech directors in the United States in mid-September and agreed to meet again in Munich in December to "explore possible cooperative efforts in standards, university programmes, equipment and materials programmes" and other areas. But despite these overtures, the question of foreign participation may run into obstacles in the US Congress, which would have to agree to change the rules by which Sematech is constituted if European participation is to be permitted.

Steven Dickman

CHANGE OF POST

New direction at ESA

Paris

REIMAR Lüst, in his fifth year as director-general of the European Space Agency (ESA), will move to Hamburg next autumn to be director of meteorology at the Max Planck Institute there.

The move had been expected: the director-general of ESA is elected for a four-year term, but Lüst had been asked to remain for another two years to avoid a change of leadership during important negotiations at The Hague in 1987.

The list of candidates for the next director-general will be closed at the end of this month, and Lüst's successor will be chosen at the ESA council meeting next March. An ESA spokesman would not speculate on the favoured candidate, but said that "it will not be a German" — successive director-generals have so far never come from the same member state. Meanwhile, Lüst has taken up another function as president of the West German Humboltstiftung. Peter Coles

US SCIENCE EDUCATION

Actions, not words, by 2007

Berkeley

DECLARING a state of emergency in science reminiscent of the days of Sputnik, a group of some 250 notables met here last week at the urging of Energy Secretary James Watkins and Nobel laureate Glenn Seaborg to find ways of improving the much-lamented state of mathematics and science education in the United States.

Among the actions called for were a mobilization of the country's many national laboratories, a nationwide teacher training programme and even a musical video to elevate mathematics and science in the minds of American elementary and high school students.

The Maths/Science Education Action Conference, convened at the Lawrence Hall of Science (LHS) in Berkeley from 9 to 11 October, was attended by representatives from education, industry, the National Science Foundation and a host of other government agencies. Its purpose was to identify specific programmes the government can adopt to reverse a steady decline in science and mathematics education.

This decline, demonstrated in several recent studies that show US students lagging behind their counterparts in many developed countries, is widely seen as threatening the technological edge of the United States. "No more talking . . . no more crying about the problem", proclaimed co-host and LHS director Marjorie Gardner. "Let's get on to the solutions. Science is a core subject, not an elective. Science is for everyone."

Despite such an admonishment, the conference was not wholly free of traditional platitudinizing — some speakers espoused such vacuous goals as making US mathematics and science students the best in the world. Nevertheless, some specific recommendations did emerge, which addressed seven areas of concern: teachers; curriculum; underrepresented students; the science student pipeline; inner-city/rural schools; national maths/science literacy; community involvement and professional involvement.

A final report is not yet available, but a draft summarizing the working group proposals argued for the development, within a year, of a streamlined process for awarding credentials to prospective teachers and a four-week training programme to reach ten per cent of existing mathematics and science teachers.

One of the conference's chief goals was to find ways to reach women and inner city and minority youth, who are traditionally underrepresented in science. It was hoped that, by establishing programmes at various national laboratories, the Department of Energy (DoE) would serve as a model for other federal agencies to help revitalize science education in targeted districts.

To that end, Watkins stopped at the Fermi National Accelerator Laboratory outside Chicago, on the way to the conference, to break ground for a $1.2 million science education centre. And Watkins told the conference that he expected to have 20 national laboratories starting similar activities within the year.

A working group chaired by Bob Goodwin, executive director of the White House Initiative on Historically Black Colleges & Universities, suggested that this effort could be expanded by bringing in private industry, colleges and state and local governments to help reach underrepresented students. The group proposed identifying ten school districts, and initiating programmes at half of them within the year. Another group suggested the music video as a way to reach those students usually left out of science and mathematics culture.

The draft summary urged the President's science adviser to establish and chair an interagency committee that would extend the DoE model by developing partnerships with other government agencies. This body would then work with the National Science Foundation (NSF) and the Department of Education on a national mathematics and science curriculum. The science adviser was also asked to establish alliances between the government and ten per cent of the country's school districts in the next year.

Those present clearly appreciated that a cabinet member had taken time to be the host of such a meeting: Watkins received a standing ovation near the end of the tiring final day. For his part, Watkins praised the reports of the working groups, saying that the bulk of them was "so much better than what we have, that we have to get a document out". Watkins also assured the meeting that its suggestions would be incorporated into DOE's National Energy Strategy, a draft of which is to be submitted to President Bush on 1 April 1990.

Although many of these ideas are aimed at producing short-term benefits, the goal of the conference was to enable the class of 2007 (children born this year) to acquire the knowledge necessary to build the country's technical and scientific base. This date is symbolic because it marks the 50th anniversary of the launching of the Soviet satellite Sputnik, an event important to conference organizers.

"We're facing many of the same problems in science and mathematics education that we faced then", said Seaborg, citing poor teacher training, an outmoded curriculum and general public illiteracy on science issues.

To put into practice the meeting's broad proposals will mean reorganizing and refocusing the already substantial resources committed to science and mathematics education, said Watkins. It will also, he admitted, take more money. "Will I have a fight? Probably. Will I get it? Yes — because I'll go and gripe to the President and he'll support me."

Robert Buderi

COMMERCIALIZATION

BTG looks towards Europe

London

THE British Technology Group (BTG), Britain's still-public technology transfer organization, is to spread its sphere of activity into mainland Europe. The group subsumes the activities of the National Research Development Corporation originally formed to patent and license the intellectual property rights (IPR) to scientific and technological discoveries made by academic researchers.

BTG has been asked by the European Institute of Technology (EIT), an international consortium supported by IBM Europe, AT&T, Glaxo, Enichem and Kodak, to deal also with innovations arising from the scientific programmes it supports at European universities. John Morton, company secretary of BTG, said EIT had recognized BTG's position as "market leader in IPR management".

EIT's objective is to stimulate cooperation between industry and universities by sponsoring research. In its first year it received nearly 1,000 research proposals, of which it supported 35, with funds ranging from 10,000 to 15,000 ECUs (1 European Currency Unit = £0.69). In phase II of the 1989 programme, nine projects will be selected, and will receive up to 750,000 ECUs each. BTG will be responsible for all the patenting and licensing of any technology in which the member companies of EIT decide they are not interested. Among other things, it will fall to BTG to decide how academics should be rewarded for their work.

Historically BTG has worked almost exclusively on British inventions, but it has lately realized that many overseas companies may have portfolios with under-developed inventions which BTG's expertise in technology and intellectual property protection skills could turn into money-spinners. BTG has strong links with the United States, and recent deals have established a firm base in Japan. The move onto the European mainland is the next stage of BTG's international expansion, and BTG has identified medium and small European enterprises as a potential growth area.

Ben Webb

CHERNOBYL CLEAN-UP

Resettlement demands

Moscow

TENS of thousands of people carried banners through the streets of Minsk, the capital city of Byelorussia, on 30 September to protest against delays and misinformation over the clean-up of contamination from the Chernobyl nuclear reactor explosion of April 1986. Although city authorities did not give official approval, the three-hour rally, attended by seven people's deputies of the USSR (members of parliament), went smoothly.

Ivan Lishtvan, vice-president of the Byelorussian Academy of Sciences, told the crowd that in Byelorussia, the republic worst affected by the Chernobyl disaster, the "optimistic forecasts" of the All-Union Committee for Radiation Defence, which has been directing the official clean-up, have proved "fallacious". He said that although the number of people affected by various illnesses is still growing, and although little is known about the effects of low-levels of radiation, the USSR Ministry of Health insists that people can live in contaminated areas.

"They want to make people live where they shouldn't", Lishtvan said, declaring that in the opinion of the Byelorussian Academy, people living in affected areas should be resettled without delay. The concerns of the academy have been echoed by the local population, which wants the federal clean-up programme in Byelorussia to heed the Academy's advice. In a few weeks, a new plan will be brought before the Byelorussian Supreme Soviet.

Academician Yevgeni Velikhov, a vice-president of the All-Union academy as well as director of the Kurchatov Institut and a people's deputy of the USSR, spoke at the rally to declare the support of the Union Academy of Sciences for the Byelorussian view.

Although there is now a policy of distributing personal dosimeters to the populace, the technical problem of making enough has yet to be addressed.

Some 1,500 dosimeters will have been produced by the end of the year, Velikhow reported, and over 100,000 in 1990. But this may not be enough to give a dosimeter to everyone who wants one, given that few people in Byelorussia are prepared to accept official statements on the extent of contamination.

Stanislav Shushkevich, pro-rector of the Byelorussian University and also a people's deputy, complained that from the outset there had been little truthful information. A demand was made for criminal charges, for misuse of office, to be laid against certain former and current officials said to be responsible for the atmosphere of secrecy around the Chernobyl explosion and its aftermath. "This will be done, sooner or later. There is no escaping an ecological Nuremberg", said author and people's deputy Ales Adamovich, who charged that because the magnitude of the disaster had been kept concealed, Byelorussia had been allocated insufficient federal money for the clean-up. And some of those resources that had been allocated were wasted, for example in the construction of new villages which will have to be evacuated anyway.

The relief fund for the affected areas of Byelorussia is now getting numerous donations, but is is not clear who will control the money, and how it will be distributed. *Izvestia* has recently reported that 542 million roubles from the federal charity fund, donated shortly after the Chernobyl explosion, was used to cover state expenses at the power station. To avoid similar misuses, speakers at the rally agreed that people's deputies for the affected areas of Byelorussia should have control of the donated funds. Whether the Supreme Soviet of Byelorussia listens to its deputies and its electorate will emerge later this month.

Igor Germenchuk/ *Novosti, Moscow*

SCIENCE IN SCHOOLS

No more creationism

San Francisco

THE state of California, which accounts for about 11 per cent of the US school textbook market, has given publishers a broad hint about the wares they should be offering. Last week, its influential Curriculum Commission voted overwhelmingly to give evolution a central place in its science framework, and to relegate creation theory to a religious setting. After a public hearing on 25 October, the state Board of Education is expected to vote on the framework in November. Textbook guidelines are revised every seven years in California, and the new recommendations will take effect in 1992. California is closely watched by publishers, who economize by producing standardized textbooks.

"Creation science is not a science and has no place in the science curriculum", said Curriculum Commission chairman Dan Chernow, a former member of the Board of Education. Creationism, he added, should be discussed along with other religious issues — either in the history and social science curriculum or the English and language arts curriculum. Robert Buderi

US AGRICULTURAL RESEARCH

Call for more spending

Washington

THE US National Research Council (NRC), the research arm of the national academies of science, medicine and engineering, is calling for an increase of $500 million in federal spending on competitive grants for research into agriculture, food and the environment.

This ambitious proposal from the NRC's board of agriculture appeared almost on the eve of President George Bush's expected announcement of across-the-board cuts in federal spending in order to meet deficit reduction targets.

But instead of watering down its recommendations in the light of the current financial climate, the NRC report, published on 6 October, argues that a substantial investment in agricultural research will improve international competitiveness and help reduce the budget deficit. The increase should be made next year and should not come from the reallocation of existing research programmes, it says.

NRC complains that there has been "little, if any, real growth" in agricultural research spending since 1955. Last year, the US Department of Agriculture's (USDA's) total research spending was only 4.6 per cent of the total federal research budget. While problems in agriculture increase, the report says, opportunities are missed: the slow application of biotechnology to agriculture is cited as a notable problem.

NRC proposes that a new independent office be set up within USDA's office of science and education to administer new grants, and that six programme areas should be established: plant systems; animal systems; nutrition, food quality and health; natural resources and the environment; engineering, products and processes; and markets, trade and policy. At present USDA spends only $40 million on competitive research grants, and in a narrower field of research.

NRC also recommends that USDA should increase the duration of grants from the present two years, as well as their size, which now averages only $50,000 a year. By contrast, the average grant awarded by the National Institutes of Health is $155,000.

The aim of the suggested programme would be to secure the proven high economic return on investment in agricultural research. Other objectives would be to attract the best talent from the entire research sector into agricultural research, to expand knowledge in all the disciplines underpinning agriculture as well as contributing to advances in other areas.

Officials in the Department of Agriculture support the proposal, but its fate will be not be known until the president submits his 1991 budget proposal to Congress. **Christine McGourty**

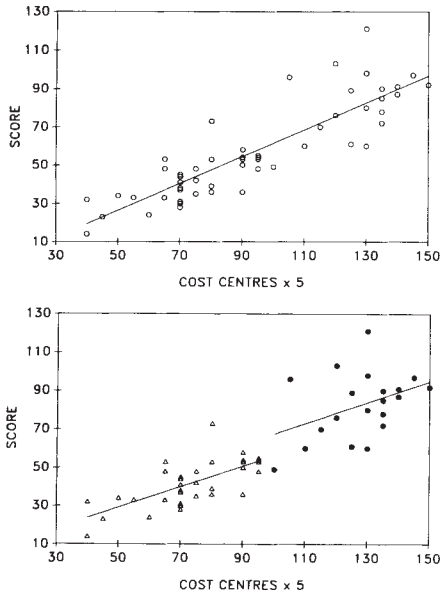
UK university rankings

SIR—The recent university research rankings by the Universities Funding Council have been based upon percentage scores, in which each university's actual score is expressed as a percentage of the maximum score the university would have obtained had each of its "cost centres" scored a 5. To get a fuller picture, it is useful to plot the published data graphically. The upper graph in the figure shows each university as a point located by its actual score (vertical axis) and its potential score (horizontal axis). Such plots are frequently used in assessing biometric data, for example in examining the relation between the size of an animal's brain and its overall body weight.

The graph shows, unsurprisingly, a high correlation ($r = 0.87$) between actual and potential scores. In other words, large institutions tend to have larger scores than smaller ones. Some universities lie conspicuously above the regression line: these are the ones picked out in newspapers as "the best". Others fall below the regression line: these are the "worst".

There is a visual impression from the graph that the universities may fall into two groupings (lower graph): a group of larger institutions (cost centres $\times 5 > 100$) where the correlation between score and size is relatively weak ($r = 0.40$), and a group of smaller ones where the correlation is stronger ($r \times 0.69$). The group of larger universities has scores greater than those expected from their size alone, as indicated by the slightly elevated position of their regression line.

It can, however, be a serious conceptual error to take position relative to the regression line as the sole measure of merit. For example, if we do the same with the size of an animal's eye relative to its total size, we arrive at the incorrect conclusion that the eye of a sparrow is better than the eye of a horse. In fact, the best single indicator of the performance of an eye is its absolute, not its relative size. A small eye is adversely affected by diffraction and by poor magnification, no matter how minuscule the size of head in which it is lodged.



Top, universities' actual score (*y* axis), compared to potential score (*x* axis), bottom.

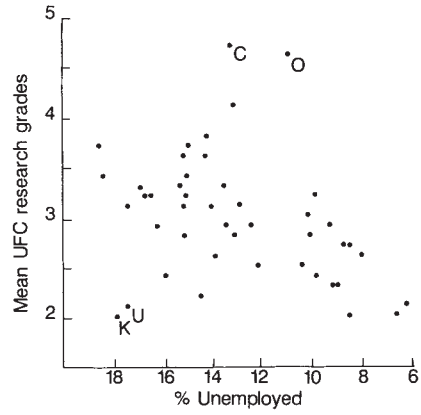
Something similar, it seems to me, might be true of universities. Reliance on percentage scores will put pressure on universities to close weaker departments and thus reduce their diversity. This could be disastrous, particularly in the case of universities that are already small. Deciding that particular universities should 'specialize' in key areas simply freezes the intellectual *status quo*. Who is to say that tomorrow's electrical engineers will not need to interact with tomorrow's geographers? The best indicator of merit (if we want one) ought to combine absolute size and relative performance. As an example, one could add the rank order of size to the rank order of distance above the regression line. Taking into account only the percentage score leads to the absurd conclusion that a university can somehow improve by closing down its weaker departments, or by amalgamating them all into a cost centre for epistemological investigations. A university that teaches no foreign languages, or which has no statistics department, might then come out as 'better' than one that teaches a reputable range of subjects. No doubt this is what the government is seeking to bring about, with the able assistance of the chief executive of the council: but there is no reason why academics should connive in destroying their own institutions.

MICHAEL J. MORGAN

28 Queens Avenue,
Muswell Hill,
London N10 3NR, UK

SIR—In the UK university system, activity in research is widely held to nurture good teaching, but supporting evidence is hard to come by. Teaching quality is difficult to assess directly but might be reflected in the employment record of graduates. In conjunction with the Universities Funding Council (UFC) research grades, employment records reported annually by the *Financial Times* provide some basis for evaluating the relationship between research and teaching performances at the level of the institution.

Far from the anticipated positive relationship between mean research grades and teaching quality, as indicated by the percentage of graduates unemployed or in short-term work, there is a negative association (see figure). The relationship is stronger in the absence of Oxford and Cambridge, which are exceptional in terms of research performance and whose graduates are moderately successful in avoiding the dole queue, and the universities of Ulster and Keele, where research and teaching both fare badly using the present criteria. No university excels in both research gradings and graduate employment, and, in general, poor research gradings are associated with enhanced graduate prospects of gaining a foothold in the world of work.



The relationship between mean research grades and percentage of graduates in 1988 unemployment or in short-term employment in 45 UK universities. Data published in the *Financial Times* on 26 August and 13 September 1989 respectively. Research grades for the colleges of the Universities of London and Wales were collated on the basis of numbers of departments assessed. Spearman's rank correlations for all universities and all universities except Oxford (O), Cambridge (C), Ulster (U) and Keele (K), are 0.400 (t=2.860, P <0.01) and 0.675 (t=5.712, P <0.001) respectively.

The 1970s and 1980s have been traumatic decades for UK universities. The view, or rather the aspiration, that research and teaching quality is uniform throughout the system is no longer tenable. There are seemingly great disparities within the system, reflecting the relative importance placed on teaching and research in individual universities. It is possible that universities that have done badly in the recent UFC grading exercise have paid the penalty of over-investment in course development and teaching. Lack of time, equipment and manpower for research and the pursuit of funds from non-UFC sources ensure low research grades. Perhaps the major error of the underachievers in research, however, is to put the interests of their students ahead of personal ambition in their chosen field of research.

W. I. MONTGOMERY

School of Biology and Biochemistry,
David Keir Building,
Queen's University of Belfast,
Belfast BT9 5AG, UK

Dog-days for the Soviet Union

Soviet scientists live more in the real world than most. The long winter that lies ahead will be particularly taxing for them. Mere organization will not help.

Moscow

THIS place is not in particularly good shape just now. The weather is neither one thing nor the other, neither mid-continentally baking hot nor arctic, and everything is a greyish-brown and damp. This is, of course, what farmers call the back-end of the year, when in the West they sell up their lifetime's plot of land and move into bungalows. In the Soviet Union, where land by definition belongs to the state, the Supreme Soviet, many people's nightly television fix has been the debate on the new law on property that, among other things, would allow land to be leased. One senses that merely to be the opening round in a five-year ideological argument. In any case, there are no bungalows into which to move.

Of the past five years of reform, only one ingredient can be counted an unmitigated success: *glasnost*. The daily and weekly newspapers and the periodicals keep turning up such interesting material that people find themselves distracted from more serious pursuits by the need to learn about the recent past. One weekly newspaper has begun to publish information about the Soviet development of nuclear weapons, complete with Beira's threat to Kurchatov that his head would be off if the first few kilograms of plutonium did not explode as promised. A principal in those events cannot accept that Beira also demanded that the estimated critical mass should be checked "with our man". *Novy Mir* has made a start on Solzhenzkin. But even free-floating *glasnost* has not yet seriously begun on Stalin's engineered famine in the Ukraine.

People also talk more freely of the situation in which they are. Some volunteer that all the past seventy years have been mistaken. Some say that it can only be a matter of time before communism is abandoned, but go on to say, "You realize I couldn't have been talking this way two years ago." Their worry, which they hide quite well, is that they may not be able to talk as freely two years from now, either.

This, it may be worth remarking, is not a selected but a general impression. Most people, when asked, answer along these lines, but usually with some degree of caution. One particularly clever man, having marshalled all the arguments why the system must be radically changed, denied that he meant that communism is dead.

Another disconcerting feature of *glasnost* is the licence it appears to have generated. Taxi-drivers habitually ask foreigners whether they have dollars to sell. (They usually offer 5 roubles to the dollar, well below the supposed black market rate of 10 or more.) Crowds of men outside hotel restaurants do the same. They seem curiously unafraid of the consequences, which on paper are horrendous; one must assume that they have squared their protectors.

It seems also to be well-established that all the newly-created cooperatives intended two years ago as the cutting edge of Mr Mikhail Gorbachev's economic reforms are not always as white as driven snow. The principle, of course, is that a cooperative can be a business in some kind of Western sense, charging what prices seem economic, but that its members must share equally in its profits. There is one account of a bookselling cooperative that bought stocks of best-selling books from a state-owned bookshop, and then sold them at more than twice the state-fixed price. It is supposed that the cooperative fixed the bookshop manager. It is no wonder that Mr Gorbachev turned last week's attack on the cooperatives in the Supreme Soviet only by referring the law concerned back to a committee.

Glasnost has also licensed magic. One television programme is said to offer healing by faith, courtesy of the public microwave channels. There is a strikingly un-Marxist credulousness in the air. Theses on the nature of the Universe (and its origin) are two a penny. Is this what the twelfth century was like?

Political parties are also springing up, but amateurishly. The founder of the Ecological Socialist Party, a mathematician turned environmental database keeper who is also a member of the real Party, was able to talk for longer about his party's objectives than about its membership (twelve) and its method of getting itself elected.

Glasnost, and compliance with the Helsinki accords, has also encouraged emigration, which is a mixed blessing. High authorities at the Soviet Academy of Sciences are sufficiently alarmed at the rate at which able people are leaving for the United States that they wonder whether a tax on overseas earnings would be equitable — and, no doubt, whether it could be administered. More poignantly, a woman friend confesses that she feels quite lonely now that her friends, mostly Jewish, are living in California and Texas.

What seems to worry people most is the problem of the republics, the partners with the Russian Republic (reaching from the Baltic at Leningrad to the Pacific) in the Union of Soviet Socialist Republics. Many people are haunted that the problem of Armenia and Azerbajan has not been dealt with.

The Armenian problem is a kind of civil war, more costly to the Soviet Union than ever Ulster has been to Britain. There is much gloomy talk about the prospect that there will be a more general civil war. One person (who himself hunts deer) says that there are too many illegal guns in people's hands. But the more common gloomy supposition of how the present confusion will be unwound is that there will be a military government for a time. If the security of the Soviet Union should be threatened, would it not be logical that the military people, even reduced in number by 400,000 as they have been, should step in?

In the circumstances, it is remarkable that there has been any progress. Yet the past two years have seen enormous change. People travel with unprecedented freedom. Almost always, they return. Laboratories are being better equipped, but exceedingly slowly, at a fraction of the rate that might allow them to "catch up". Perhaps the best sign, in the past two years, has been the appointment of really able scientists as heads of academy institutes on no other basis than merit; many of the new guard are not even members of the Party. Yet the cleverest are happiest when their laboratories are aweay from Moscow.

That, for the scientific enterprise in the Soviet Union, must be where the future lies. For better or worse, the Soviet Academy of Sciences is still well-respected and could regain much of the influence it is losing to the State Committee on Science and Technology if it were more efficient. There is endless speculation as to how the next elections, due in the first half of next year, will resdistribute people among positions, but the academy and its members have hardly begun to think of how it should itself be organized. Yet as things are, the weekly meetings of the governing council (praesidium) bring together more than 50 people from all over the Soviet Union for a day-long meeting. A little procedural *perestroika* would not come amiss.

John Maddox

PHOBOS 2 ENCOUNTER

Probing martian space . . .

Mark Saunders

OF all the planets, Mars must surely evoke the greatest sense of mystery and fascination. Excitement was high, therefore, when earlier this year the Soviet Phobos 2 spacecraft became the first probe since NASA's two Viking landers in 1976 to reach the red planet. Despite the premature failure of both its spacecraft (Phobos 1 was lost while *en route* to Mars and contact with Phobos 2 ceased after just 57 days in orbit; Sagdeev and Zakharov discuss these failures in detail in their review article[1] on page 581), the ambitious Phobos project did return new and important scientific data, as is clear from the initial results reported on pages 581–618 of this issue.

Before the Phobos mission, martian space was even less well understood than the space environments of the more distant planets Mercury, Jupiter and Saturn. The only previous missions to study aspects of the interaction of the solar wind with Mars were the Mariner 4 flyby in 1965, which at closest approach barely crossed the bow shock, and three Soviet orbiters (Mars 2, 3 and 5) in the early 1970s. The Soviet spacecraft all orbited at least 1,000 km above Mars, an altitude that is too high for critical aspects of the solar-wind interaction to be studied. Interpretation of the scant Soviet recordings thus became controversial, especially on the question of Mars's intrinsic magnetic field. So, before Phobos, it was unclear whether the solar wind at Mars was deflected totally by an ionosphere (as at Venus) or whether a magnetic field (as at Earth) also played a role.

During its 57 days in orbit Phobos 2 made made 5 elliptical orbits with a closest approach to Mars of 850 km followed by more than 100 circular orbits at an altitude of 6,000 km. But the elliptical orbits are proving the most rewarding, because they reached the lower altitudes vital for answering the key questions. Remote data concerning the surfaces of Mars and Phobos (the larger of the two martian moons after which the mission was named) are highlighted in the accompanying News and Views article by McKinnon opposite.

Although Phobos 2's orbits were not ideal for probing the plasma around Mars, initial data analyses[2-4] suggest that the character of martian space is controlled by the solar wind and its embedded interplanetary magnetic field (IMF), but there is also tentative evidence of a weak planetary field[2]. Major boundaries are similar to those found at Venus, although the scales are much shorter, indicating the relevance of microscopic plasma effects. The outward leakage of Mars's thin atmosphere seems large enough to have influenced the planet's evolution. These findings may have wider relevance in the Solar System as similar space environments occur at Venus, at the comets and at Titan.

One important result to emerge from the new data is that the solar wind at Mars interacts dominantly with an ionosphere[2]. This is clear from the Phobos 2 passes through the martian wake, which show that at altitudes of 6,000 km, Mars has a well-developed magnetic tail consisting of two opposed bundles of magnetic flux (called lobes) whose polarities are controlled by the direction of the upstream IMF (Fig. 1). Thus it seems that most of the magnetic flux in the martian magnetotail is accreted from the solar wind in the manner Alfven[5] proposed for comet magnetic tails and which is observed at Venus. Such tail accretion is a direct consequence of the interaction of a flowing magnetized plasma with a planetary ionosphere. The magnetometer team[2] also report spectral-analysis evidence for a small internal magnetic field, but this result should be viewed with caution until substantiated further.

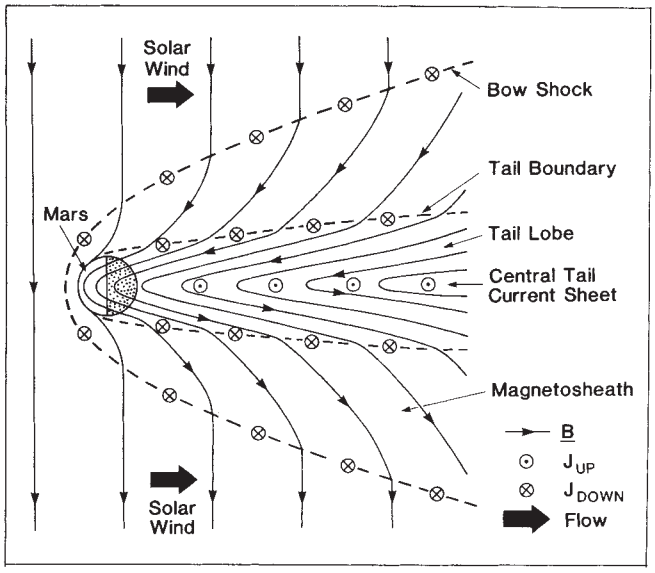
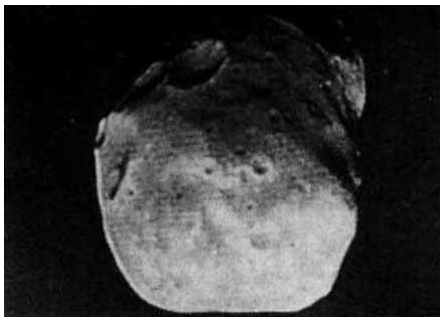


FIG. 1 Electromagnetic aspects of the interaction of the solar wind (bold arrows) with Mars, consistent with the initial Phobos 2 findings and illustrating the structure of an accreted magnetic tail. The diagram is drawn for steady east–west interplanetary magnetic field conditions and field lines, *B*, are shown projected into the plane of the figure. Circled dots and crosses indicate the direction of electric current flow *J*.



The eponymous Phobos.

Also important is the information obtained on the plasma and magnetic structure of Mars's dayside space environment, the only region probed thoroughly on previous missions being the bow shock. Figure 2 illustrates the main dayside boundaries and their locations, inferred from measurements from the elliptical orbits. The plasma boundary nearest the planet has been called the planetopause (though ionopause may be a more appropriate term). It separates the shocked solar-wind plasma in the magnetosheath from a region of purely ionospheric plasma consisting mainly of atomic oxygen O^+. The interplanetary magnetic field piles up above the planetopause (Fig. 2), but a large amount of magnetic flux penetrates this boundary because the solar wind's dynamic pressure usually exceeds the plasma pressure of Mars's ionosphere[6]. The compressional stresses exerted on these slowly moving dayside field lines will act to squeeze plasma tailward along the draped magnetic field. This effect may be significant for the observed[3,4] downtail escape of ionospheric plasma.

A further source of tailward plasma loss is the accelerated field-aligned beams of O^+, with energies of up to several kiloelectron volts, found[3,4] near the central tail current sheet. The origin of these beams remains uncertain. One suggestion is that they arise from the same acceleration process occurring above the Earth's auroral oval; however, Phobos 2 did not detect aurorae at Mars. An alternative origin may involve current sheet acceleration associated with the hairpin bend of the field lines in the central tail (Fig. 1).

The tailwards escape of ionospheric plasma in the martian wake is estimated to be between 1 and 2 kilograms per second[3,4]. This rate may seem small but, if maintained, it would result in Mars's thin atmosphere being lost within 10^9 years, which is less than the age of the Solar System. Assuming that the atmosphere was replenished through the evaporation of surface water or ice frozen into the martian soil, then the weakness (or absence) of an intrinsic magnetic field may well have contributed to the loss of

both the martian atmosphere and hydrosphere through the direct impact of the magnetized solar wind on the ionosphere.

There is a special aspect of the martian system that makes its plasma environment particularly challenging to understand. Whereas the main features of other planetary environments are described well by the equations of magnetohydrodynamics, in which the motion of all particles is ascribed to a fluid flow, the observed scale sizes at Mars are comparable to the ions' local radii of gyration in the magnetic field (1,000 km for a solar-wind ion). Thus the solar wind's interaction with Mars is modified by effects due to such finite Larmor radii[7]. One clear example of such a microphysics effect is the width of the foot boundary associated with the bow shock[2,8]. As is evident in Fig. 2, the shock foot (a shock precursor one ion-gyroradius wide)

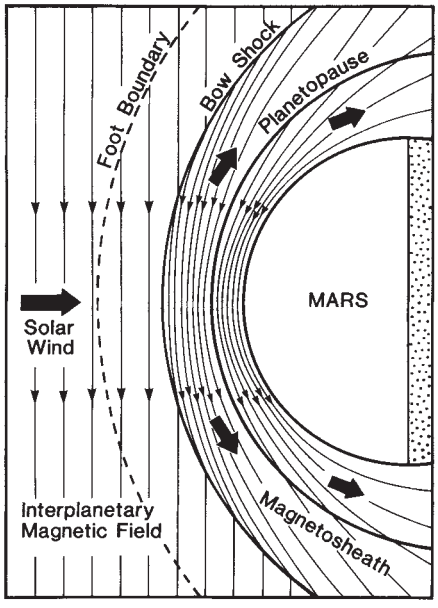


FIG. 2 The dominant dayside features of the solar-wind interactions with Mars seen by Phobos 2. The diagram emphasizes the main boundaries, the magnetic field geometry (arrowed lines) and the plasma flow (solid arrows).

is now comparable in width to the magnetosheath and is not an insignificant feature as it is at Earth.

Further important findings will follow these early analyses, but it is doubtful that Phobos 2 can answer the crucial question of whether Mars has a small or zero intrinsic magnetic field, because the 850 km periapsis (rather than the nominal 500 km) was too high to probe the ionospheric regions where an internal field would be most detectable. Indeed a periapsis as low as 100–200 km may be required for a definitive answer, and this will have to await the planned NASA Mars Observer and Soviet MARS 94 missions. □

Mark Saunders is in the Blackett Laboratory, Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK.

. . . and martian surfaces

William B. McKinnon

THE loss of Phobos 2 in Mars orbit is a tragedy, of course, but it is not the end of the story. Although it is not widely recognized, many of the experiments on board were intended for Mars as well as Phobos, and it is probably these experiments that yielded the most new data and that will, eventually, provide the most new scientific understanding. The thermoscan (thermal infrared) images, in particular, are spectacular[9]. They instantly show much of the information previously painstakingly obtained by the Viking IRTM (infrared thermal mapper), and at better spatial resolution. And they allow one quickly to note the relatively warm and cool areas, interpret them in terms of albedos and thermal inertias (that is, soil-like or rock-like thermal properties), and correlate all this with local geology.

The long history of Viking IRTM analyses shows that many factors, such as surface pressure, must be taken into account for proper interpretation, so it is premature to expect detailed geological interpretation at this time. Nevertheless, analysis of thermoscan images can draw on the IRTM experience. For example, high-resolution but broadband thermoscan radiometer data can be combined with lower resolution but multispectral IRTM data for maximum effect.

Phobos 2 also carried an infrared imaging spectrometer (ISM), and was the first spacecraft to do so at another planet. Several hundred spectra of Phobos were obtained along with some tens of thousands of spectra of Mars. The martian spectra sampled most of the principal geological terrains on the planet, with the unfortunate exception of the polar caps. In principle, it should be possible to use the details of the spectra collected to identify or place constraints on the types or amounts of various minerals in the martian soil and bedrock (if exposed). The example in the paper on the ISM results[10] is the relative depth of the water hydration band near 3 μm. The authors are careful not to overinterpret the results, as the depth of an absorption feature does not uniquely indicate the relative amount of the absorber: grain size, the presence of opaque (and hence masking) minerals, degree of crystallinity and even temperature affect spectral band depths.

Refined analyses exploiting the results of years of ground-based spectroscopy should provide more secure identification of primary igneous minerals, and most importantly, perhaps shed some light on the controversy surrounding identification of carbonate-bearing minerals in the martian soil. Carbonate is, naturally enough, a sink for atmospheric CO_2, and its presence may indicate a more aqueous (and clement) Mars in the past.

The spacecraft brought a multichannel infrared radiometer to Mars as well, but the results presented so far concern only Phobos[11]. The data indicate intriguing inhomogeneities in the inner moon's structure and composition. ISM data for Phobos also indicate variations in hydration and silicate spectral features[10]. The 3-μm hydration absorption feature is markedly weaker than for Mars, which may at first seem surprising given that Phobos (and Deimos) are often viewed as captured carbonaceous asteroids and that carbonaceous meteorites are mostly water-rich clay minerals. However, the surfaces of many carbonaceous asteroids, particularly those in the outer asteroid belt, are rather dry, and ground-based spectroscopy has recently shown that Deimos possesses at most only a small absorption feature at 3-μm, implying that it is relatively dry[12]. The ground-based experiment is too difficult to do for Phobos because of scattered light from Mars, but the telescopic and ISM data combined may indicate that Phobos and Deimos are similar carbonaceous bodies with closer to anhydrous surfaces.

It is also interesting to note that the Phobos mission indicates that the moon's density is about 1.95 g cm^{-3}, which is somewhat less than that determined by Viking. The imaging team reports that this is consistent with substantial internal porosity for a rock body[13], which is true, but some noted astronomers still favour an icy interior. Paradoxically, an icy interior would not conflict with a dry surface, because this is what is expected of outer belt asteroids according to current thinking[14]. The prospect of icy moons circling a thirstless desert below is only one of many martian mysteries to be explored by further analysis of Phobos 2 data. □

William B. McKinnon is in the Department of Earth and Planetary Sciences and McDonnell Center for Space Sciences, Washington University, Saint Louis, Missouri 63130, USA.

1. Sagdeev, R.Z. & Zakharov, A.V. *Nature* **341**, 581–585 (1989).
2. Riedler, W. *et al. Nature* **341**, 604–607 (1989).
3. Lundin, R. *et al. Nature* **341**, 609–612 (1989).
4. Rosenbauer, H. *et al. Nature* **341**, 612–614 (1989).
5. Alfven, H. *Tellus* **9**, 92–96 (1957).
6. Luhmann, J.G. *et al. J. geophys. Res.* **92**, 8545 (1987).
7. Moses, S.L. *et al. Geophys. Res. Lett.* **15**, 429 (1988).
8. Grard, R. *et al. Nature* **341**, 607–609 (1989).
9. Selivanov, A.S. *et al. Nature* **341**, 593–595 (1989).
10. Bibring, J.-P. *et al. Nature* **341**, 591–593 (1989).
11. Ksanfomality, L.V. *et al. Nature* **341**, 588–591 (1989).
12. Bell, J.F., Piscitelli, J.R. & Lebofsky, L.A. *Lunar planet. Sci.* **XX**, 58–59 (1989).
13. Avanesov, G.A. *et al. Nature* **341**, 585–587 (1989).
14. Lebofsky, L.A. *et al.* in *Origin and Evolution of Planet and Satellite Atmospheres* (eds Atreya, S.K., *et al.*) 192–229 (University of Arizona Press, Tucson, 1989).

AIDS

Complexities and strategies

Peter Newmark

SET the difficult task of deciding the best research strategies in AIDS, participants at a recent meeting* were somewhat diverted, but not particularly cheered, by the recent US decision in favour of giving AZT, still the only licensed drug against HIV (human immunodeficiency virus), to those who are HIV-infected but have no symptoms of disease. Several speakers voiced the opinion that the decision was made before it was certain that AZT at an early stage does more than delay progression, in which case it might be wiser to hold back on treatment for more advanced stages of the disease. With the US data not available for scrutiny, N. Clumeck (St Pierre University Hospital, Brussels) announced that equivalent European trials would be continued for the time being.

In any case, AZT is orders of magnitude too expensive to be used in developing countries, which account for about half of the 5–6 million people in the world who are estimated already to be infected with HIV (J. Chin, WHO), although only the tip of that iceberg is so far visible, in the form of a cumulative total of 180,000 cases of AIDS.

As G. Noble (Centers for Disease Control, Atlanta) put it, trying to monitor trends in HIV prediction by extrapolating backwards from HIV cases is becoming extremely antiquated, and will be even less adequate as drugs such as AZT lengthen the time from infection to AIDS. Thus, there is a pressing need for widespread surveys of rates and trends in infection. The ideal system would be random door-to-door surveying, but this is controversial, expensive and subject to an excess of refusals among those most at risk. Testing of anonymous blood samples is better but leaves unanswered the question of the likely cause of infection. What is often lacking, concluded many of the delegates to the conference, is the political leadership necessary to bring about and finance adequate surveys.

Equally lacking is the leadership necessary to promote measures that would reduce the rate of infection of those most at risk. An example of what can be achieved is the reduction in annual infection rates of a cohort of San Francisco homosexuals from 20 per cent in 1982 to 1–3 per cent in the past three years. This is attributed largely to behavioural changes, although if more susceptible individuals were the bulk of the 50 per cent of the cohort that had already become infected by 1982, the annual rate would be bound to have fallen to some extent thereafter even without behavioural change. (Current estimates are that AIDS will kill 4 per cent of the population of San Francisco.)

Another example is the fall in the annual infection rate of intravenous drug users in San Francisco in the past two years, associated with a programme of education, the provision of bleach as a needle disinfectant and (in the face of California state law) the promotion of a scheme of needle exchange (D. Francis, Dept. Health Sciences, Berkeley). It was generally concluded that attempts to promote behavioural change work best at a local level based on an adequate understanding of local high-risk practices.

Animal models

Unfortunately, monkeys are still the best animal models, but even they have limitations. More than six years after being innoculated with human brain tissue containing HIV-1 (autopsy samples from what are now known to have been cases of AIDS dementia), chimpanzees show no sign of disease (C. Gajdusek, National Institutes for Health). Lymphadenopathy follows the infection of rhesus monkeys with HIV-2 (G. Hunsmann, German Primate Centre, Gottingen), but after several months no other signs of AIDS have appeared. This is perhaps inevitable, given that a study of Senegalese prostitutes infected over three years ago with HIV-2 has shown it to be far less pathogenic than HIV-1 (M. Essex, Harvard University). So, by fairly general agreement, the best system at present remains the AIDS-like disease that follows infection of rhesus monkeys with simian immunodeficiency virus (SIV) from other species. If the molecularly cloned SIV that kills monkeys in as few as ten days after infection (J. Mullins, Stanford University) is shown to be a viable model for the very much slower HIV-induced diseases, progress should be much speeded.

A potential substitute for the monkey is the cat, which develops an AIDS-like disease when infected with the recently discovered feline immunodeficiency virus. However, its complete sequence shows it to be only a distant relative of HIV (and all other lentiviruses) and epidemiological evidence suggests that it can be transmitted by close contact, unlike HIV (E. Sparger, University of California, Davis). An even more recent rival is the SCID-hu mouse — a genetically immunodeficient mouse with circulating human T cells derived from implanted human fetal liver and thymus. If HIV is given in the thymus or intravenously, it infects thymus cells and replicates there; AZT will inhibit but not completely prevent the infection (J. McCune, SyStemix, Palo Alto). It has yet to be shown, however, that SCID-hu mice can mount an immune response against HIV, so the model is not yet, and may never be, useful for vaccine testing.

Vaccines

The extent of the immune response evoked in 10 volunteers given a vaccine comprising the envelope protein of HIV, made in yeast, with a synthetic muramyl tripeptide adjuvant was reported by D. Dina (Chiron). In most cases, antibodies were produced but the response was neither sustained, nor did it include neutralizing antibodies, which are considered to be essential for protection. It is both on this kind of outcome and on theoretical grounds that the tide of the meeting turned against simple subunit vaccines. At the other extreme is the possibility of using killed-virus vaccines. Delegates were informed, unfortunately second-hand, of the apparently successful protection of a number of rhesus monkeys by a killed-SIV vaccine.

It seems unlikely, however, that killed-HIV vaccines will ever be considered safe enough for human use. Instead, a strong case was made for focusing efforts on live vector viruses, such as vaccinia virus or adenovirus engineered to contain appropriate HIV genes (G. Ada, Johns Hopkins University). These would have the great advantage over subunit vaccines of being able to generate an immune response against cells containing HIV, which is probably the predominant form in which the virus is transmitted. An unsophisticated live vector vaccine would contain, for example, the whole HIV envelope gene; more sophisticated versions would express only those parts that generate strong immunity, including the 9-amino-acid peptide identified by D. Bolognesi (Duke University) as the principal neutralizing epitope on the envelope protein[1]. Antibodies against this epitope do not

A PROBLEM in comparing AIDS research from different laboratories is the lack of standard reagents, including viruses. This is particularly the case in research on vaccines. Help is now at hand in Europe in the form of a new European Communities initiative, known as EVA (European vaccine against AIDS).

Most of EVA's total budget of ECU 2.5 million (£1.8 million) will be used on the production and distribution of reagents. They will be produced in a number of laboratories to agreed specifications. Quality control, storage and distribution will be looked after by the National Institute for Biological Standards and Control in South Mimms, UK, which is open to suggestions as to which reagents are most needed until the end of this month. □

* Institut de la Vie symposium, *Biomedical Research Strategy on AIDS*, Yverdon-les-Bains, 24–28 September 1989.

block binding of the protein to its receptor, the CD4 molecule, but prevent fusion of the virus with the cell thereafter.

It is most unlikely that a single epitope will provide sufficiently broad protection against the great variety of strains of HIV. And it was generally agreed that to produce broad T-cell-mediated immunity will require several epitopes. Some, but not much, concern was expressed that the ability of certain antibodies to enhance, rather than prevent, the entry of HIV into cells (J. Levy, University of California, San Francisco) might compromise the use of vaccines.

Another matter of concern, not only to the development of vaccines but also to the understanding of AIDS, emerged from suggestions that the HIV/SIV isolates on which most studies are based may not be representative of disease-causing forms. For example, the SIV with a truncated envelope protein that appears when isolates from macaque monkeys are cultured in the laboratory is the result of adaptation to growth in human cells rather than the predominant form *in vivo* (Mullins)[2].

More dramatically, the most frequently represented variants among the population of HIV molecules found after culturing an isolate are usually only minor components in the uncultured population, sampled directly by polymerase-chain-reaction technology, at least in terms of their *tat* gene sequences (S. Wain-Hobson, Pasteur Institute)[3]. Further studies along these lines should help establish not only whether the most pathogenic variants of HIV and SIV require the help of other variants in order to replicate, as some researchers believe, but also the functional importance of the apparent evolution of HIV variation in individuals during progression towards AIDS (Wain-Hobson; E. Fenyo, Karolinska Institute).

Although it remains unclear exactly how HIV causes disease, and why there is a long period of latency, several new facts have come to light. First, it now seems that the virus is present in a much greater proportion of peripheral blood cells than has been thought, although still only one in every hundred or few hundred cells is infected even in AIDS (D. Ho, University of California, Los Angeles; see also ref. 4). Second, there is increasing evidence that CD8 lymphocytes suppress HIV replication in CD4 cells, perhaps by means of a lymphokine (Levy). Finally, the product of the *tat* gene, secreted from HIV-infected cells, may be one of the factors promoting the development of Kaposi's sarcoma (R. Gallo, National Institutes for Health).

Whereas those at the meeting charged with trying to reach a consensus view on the best strategies to follow were able to select some areas on which to focus, new information on the complexities of HIV and the diseases it causes made it clear how foolhardy it would be to narrow the focus more than a little at present. □

Peter Newmark is Deputy Editor of Nature.

1. Javaherian, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 6768–6772 (1989).
2. Hirsch, V. *et al. Nature* **341**, 573–574 (1989).
3. Meyerhans, A. *et al. Cell* **58**, 901–910 (1989).
4. Schittman, S.M. *et al. Science* **245**, 305–308 (1989).

QUANTUM TUNNELLING

Barrier traversal time

Rolf Landauer

CONSIDER a large box with a particle that can escape through a thin tube. The particle spends a long time bouncing around the box until it escapes and then a shorter time in the tube, while escaping. A similar distinction in timescale exists for quantum mechanical tunnelling out of a trap through a barrier, although this has received limited recognition. The time taken by the final escape event has now been measured in a subtle and definitive experiment by a group at Saclay[1].

In classical mechanics, a particle moving sufficiently far against an opposing force field will eventually have its velocity reduced to zero, at which point it will turn around. In quantum mechanics, a particle has some probability of going further and penetrating a barrier or wall, which would certainly stop the classical particle. This tunnelling has been observed and discussed in many contexts such as α-particle emission from nuclei, fusion catalysed by muons, the Esaki diode and the scanning tunnelling microscope. It is frequently maintained that the final tunnelling escape through the barrier takes zero time[2], or that quantum mechanics does not allow us to discuss the timing. The latter view, typically held in casual discussions, is based on a presumption of excessively simple methods of measuring the barrier traversal time.

In the past decade a degree of consensus has emerged in the theory of this barrier traversal time (see, for example, refs 3–7). Although each contributor has a variation of a related viewpoint, the papers all emphasize an expression in which the usual imaginary sign of the momentum in tunnelling, $p=\sqrt{2m(E-U)}$ (U the potential energy exceeding E, the total energy), is disregarded, and the magnitude of that momentum, divided by the mass m, is regarded as a barrier crossing velocity. This barrier traversal time is familiar from instanton theory approaches to tunnelling, but there it is an imaginary time, without physical intepretation. Some controversy continues, for example between Büttiker and myself[8] and Lowe and Collins[9], and between Hauge *et al.*[10] and Leavens and Aers[11].

It has been argued that this expression for the traversal time was confirmed by an experiment[12] in which a tunnelling current was reduced by the presence of a deflecting magnetic field. But this result can be interpreted equally well by direct appeal to the time-independent Schrödinger equation, without use of the traversal time[13]. A more dynamic form of experimental confirmation is desirable. Three such experiments have been proposed, including one which has yielded only preliminary results[14]. A second[15,16] uses a configuration in which, during tunnelling, the electron is attracted to an electrostatic image charge on the electrode which it is leaving. The extent to which this image charge has time to spread, while the electron is tunnelling, is used as a clock for measuring the tunnelling time. The third experiment by Esteve *et al.*[1] has now yielded very clear results. More important than the exact result and its relation to theoretical controversies, is the fact that a timescale associated with the barrier traversal can be measured, and is a real (not imaginary) quantity. The occasional advocacy of complex traversal time is puzzling: we have yet to see a stopwatch with complex numbers on its dial.

Esteve *et al.* use a Josephson junction (an electrical insulator sandwiched by two superconductors) supplied with a constant

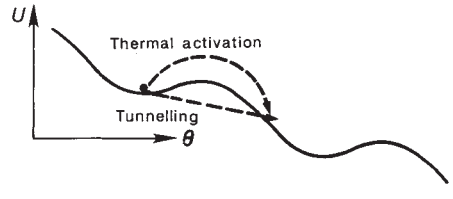


FIG. 1 Metastable states of a circuit containing a Josephson junction. Movement along the abscissa, the phase difference across the junction, generates a voltage V: $d\theta/dt = 2eV/\hbar$. The energy U depends on θ and the (constant) current J through the junction according to $U = (J_0\cos\theta - J\theta)h/2e$, where J_0 is the maximum current the junction can carry without a voltage. J_0 determines the height of the undulations, and J, the tilt of the 'washboard' potential. With $J<J_0$ the potential exhibits pockets of local stability; the circuit can remain in such a metastable state for some time (akin to the particle of the first paragraph rattling in a box) before tunnelling out. Resistive dissipation during the tunnelling lowers the energy of the system (the broken arrow indicating tunnelling is not horizontal). The value of the effective resistance depends on whether the barrier traversal time is larger or smaller than the characteristic timescales of the rest of the circuit, and also affects the probability (rate) of tunnelling escapes.

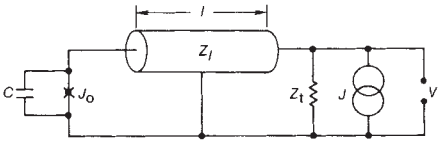


FIG. 2 Circuitry of the experiment by Esteve *et al.*[1]. The Josephson junction (✖) passes a current J, less than its critical current J_0, from a constant source; it has a capacitance, C. It is also connected to a 'terminating' resistor, Z_t, through a transmission line Z_l (or delay line) of variable length l (and hence variable delay). The decay of a metastable state is apparent because of the voltage V that is generated and which can be measured. After each decay event, the circuit can be reset, and the time elapsed until the next decay measured.

current. The junction permits the tunnelling of electron pairs through an insulator, with the electrons in the pair coupled by the mechanism responsible for superconductivity in the adjacent electrodes. The state of such a junction is characterized by the difference in quantum-mechanical phase across the insulator, related to the voltage V by $d\theta/dt = 2eV/\hbar$ (e is the electron charge; $\hbar$, Planck's constant; t, time). Esteve *et al.* exploit the fact that the energy of their circuit containing the junction varies with the phase: under the appropriate conditions, there is a series of states of local stability in which the circuit can remain (Fig. 1).

A circuit trapped in one of the metastable wells can escape either by thermal activation over the barrier or by tunnelling quantum mechanically through it (Fig. 1). It is the barrier traversal in the latter process, which dominates at low temperature, that Esteve *et al.* time. Their experiments are carried out at 18 millikelvin and above. In addition to the tunnelling pairs there is a normal current arising from single electrons flowing across the junction. This flow can arise from leakage or from the tunnelling of uncoupled single electrons, and also from current carried through any additional conductor 'shunting' the junction. These are all dissipative transport mechanisms, supplying viscous damping for the motion in the potential depicted in Fig. 1. The viscous forces reduce the probability (or rate) of tunnelling: roughly speaking, the circuit loses energy as it tunnels and so drops further below the barrier, which consequently becomes more impenetrable. By measuring the mean lifetime of the metastable state — simply by timing many decay events — Esteve *et al.* can measure the effective resistance during the barrier traversal and so derive, as described below, the time taken for the traversal itself.

A transmission line, of mechanically variable length and terminated by a resistor, is connected across the junction (Fig. 2). The resistance of the termination is considerably smaller than that of the transmission line, so that voltage waves or pulses propagating towards it are reflected back to the junction. As the circuit tunnels through the potential barrier, $d\theta/dt$ is positive, so that a voltage is generated (as given by the equation above) which propagates down the transmission line.

What effect does the terminating resistor have on the tunnelling event? With a sufficiently long transmission line, the tunnelling event will be concluded before the voltage wave reaches the resistor. The dissipative energy losses will be determined by the characteristic impedance of the transmission line. If the line is shorter, the voltage wave can reach the termination and be reflected back to the junction while the traversal occurs. The effective damping will, then, depend also on the resistor.

Esteve *et al.* study the cross-over between the two situations and find a traversal time whose magnitude agrees with that predicted in refs 3–7. The experiment requires sophisticated subsidiary measurements to determine the circuit parameters, and the analysis includes a full treatment[17] of the frequency dependence of the circuit's response.

In most solid-state tunnelling situations, electrons, rather than the state of a whole circuit, are involved in the tunnelling. That leads to traversal times which relate to atomic distances and to typical electron velocities. These times are of the order of 10^{-14} or 10^{-15} seconds. The actual value found by Esteve *et al.*, 78 picoseconds, is remarkably long by comparison, which is why a mechanically variable transmission line, functioning on a similar timescale, can be used in the timing. □

Rolf Landauer is at the IBM T.J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA.

1. Esteve, D., Martinis, J. M., Urbina, C., Turlot, E., Devoret, M. H., Grabert, P. & Linkwitz, S. *Physica Scr.* T**29**, 121–124 (1989).
2. Kane, E.O. in *Tunneling Phenomena in Solids* (eds Burstein, E. & Lundquist, S.) 1–11 (Plenum, New York, 1969).
3. Jonson, M. *Solid St. Commun.* **33**, 743–746 (1980).
4. Stevens, K.W.H. *J. Phys.* C**16**, 3649–3661 (1983).
5. Büttiker, M. & Landauer, R. *Phys. Rev. Lett.* **49**, 1739–1742 (1982).
6. Leavens, C.R. & Aers, G.C. *Solid St. Commun.* **63**, 1101–1105 (1987).
7. Leavens, C.R. & Aers, G.C. *Solid St. Commun.* **63**, 1107–1111 (1987).
8. Büttiker, M. & Landauer, R. *J. Phys.* C**21**, 6207 (1988).
9. Lowe, D. & Collins, S. *J. Phys.* C**21**, 6210–6213 (1988).
10. Hauge, E.H. *et al. Phys. Rev.* B**36**, 4203–4214 (1987).
11. Leavens, C.R. & Aers, G.C. *Phys. Rev.* B**39**, 1202 (1989).
12. Guéret, P. *et al. Europhys. Lett.* **3**, 367–372 (1987).
13. Eaves, L. *et al.* in *The Physics and Fabrication of Microstructures and Microdevices* (eds Kelly, M.J. & Weisbuch, C.) 343–350 (Springer, Berlin, 1986).
14. Lucas, A.A. *et al. J. Vac. Sci. Technol.* A**6**, 461 (1988).
15. Guéret, P. *et al. Appl. Phys. Lett.* **53**, 1617 (1988).
16. Guéret, P. *et al. Solid St. Commun.* **68**, 977 (1988).
17. Leggett, A.J. *Phys. Rev.* B**30**, 1208–1218 (1984).

PALAEONTOLOGY

Hearing in early mammals

Kenneth Kermack

CHANGES in the detailed anatomy of the ear are central to mammalian evolution, but most attention has been devoted to the middle ear. The inner ear is wholly enclosed within the petrosal bone of the skull and is consequently much less accessible. Serial sectioning of the petrosal is the ideal way of learning about the inner ear. Using this technique in a new study[1] of *Morganucodon watsoni*, from the early Jurassic, Graybeal and co-workers show that the hearing of this primitive mammal was much closer to that of birds and reptiles than to that of modern mammals.

In most vertebrates, the jaw hinge is formed from the quadrate in the skull and the articular in the lower jaw. In mammals, these two bones form part of the chain of three bones transmitting sound from the eardrum to the inner ear; the jaw joint is formed from two other bones, the squamosal in the skull and the dentary in the lower jaw. This unique squamosal–dentary jaw articulation is a diagnostic feature of mammals. *Morganucodon* had this joint between the squamosal and the dentary, but retained the older articulation between the quadrate and articular[2,3]: the two latter bones formed part of the jaw hinge as well as transmitting sound from the eardrum to the inner ear.

The hearing of mammals differs from that of other vertebrates in a manner that is often overlooked. The latter have an upper limit of frequency of about 10 kHz, the sensitivity of the ears even of owls declining rapidly above 12 kHz (ref. 4). Mammals are in a different league: the ears of the domestic cat are efficient at over 50 kHz, and even man, a poor performer among mammals, can better 15 kHz. This increased frequency range in modern mammals was made possible by two novel developments. The first is the separation of the mechanism for the conduction of sound from the jaw articulation: the middle-ear bones become much lighter and less constrained in mammals than in their immediate ancestors, and so conduct high-frequency sounds with greater efficiency. Second, the basilar membrane, which supports the organ of Corti (the actual sound receptor), becomes much longer. This is housed in a duct in the inner ear called the cochlea, so a long basilar membrane implies a long cochlea. In marsupials and placentals the cochlea is coiled up like a snail shell inside the petrosal. In living monotremes (the platypus and spiny echidna) the cochlea is shorter and curved rather than coiled.

Several petrosal bones have been preserved among the *Morganucodon* material from South Wales, including one

almost perfect specimen. The external features of the petrosal have been described in detail[5]. Graybeal and coworkers[1] sectioned four incomplete petrosals, traced the courses of nerves and blood vessels inside the bone and identified the openings at which these canals terminate with more certainty than is possible from inspecting the external surface of the bone alone. It is gratifying that they confirmed the existence of many of the canals and openings identified from the original description[5], although some of these had been lost on their less complete specimens. Graybeal *et al.* conclude that the relative length of the basilar membrane in *Morganucodon* resembled that of a bird or a modern reptile more closely than that of a modern mammal.

The obvious deduction from this is that the frequency response of the inner ear of *Morganucodon* was also much more like that of a reptile or a bird than a modern mammal. This was inevitable. *Morganucodon* stands right at the beginning of mammalian evolution: it is a mammal because the squamosal and dentary played a part in the jaw articulation, but the quadrate and articular had not been freed to become solely part of the sound transmitting mechanism of the middle ear, as in modern mammals. There is no use in having an inner ear capable of detecting high frequency sounds which the middle ear cannot transmit. □

Kenneth Kermack is Emeritus Professor of Vertebrate Palaeontology in the University of London, and an Honorary Research Fellow in the Department of Biology, University College, Gower Street, London WC1E 6BT, UK.

1. Graybeal, A. *et al. Zool. J. Linn. Soc.* **96**, 107–117 (1989).
2. Kühne, W.G. *Palaont. Z.* **32**, 197–235 (1958).
3. Kermack, K.A. *et al. Zool. J. Linn. Soc.* **53**, 87–175 (1973).
4. Knudsen, E.I. in *Comparative Studies of Hearing in Vertebrates* (eds Popper, A.N. & Fay, R.R.) 289–323 (Springer New York, 1980).
5. Kermack, K.A. *et al. Zool. J. Linn. Soc.* **71**, 1–158 (1981).

MITOCHONDRIAL DNA

Small, beautiful and essential

L.A. Grivell

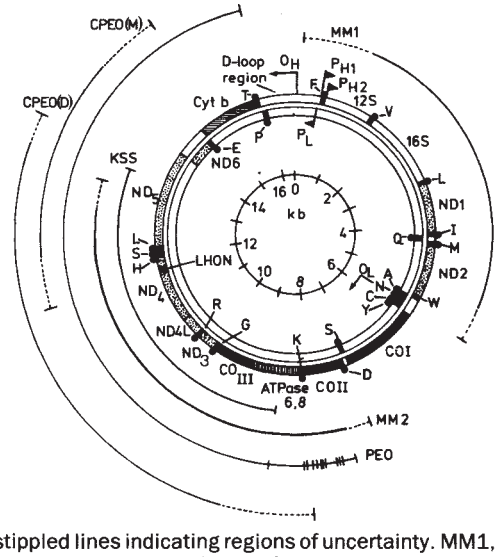
HUMAN mitochondrial DNA (mtDNA) is remarkable both for its small size and for the immaculate economy of organization of the genes encoded by it[1]. Regarded by some as merely a relic of the event that brought a symbiotic bacterium into an ancestor of the eukaryotic cell, by others as a molecular biological curiosity for the ways in which conventional mechanisms are flouted during expression of its genes, and by yet others as a convenient marker for tracing recent evolution, mtDNA is now attracting attention for another reason. As a spate of recent reports demonstrate[2-8], a number of human neuromuscular diseases have structural abnormalities of mitochondria as their common feature and these appear to be caused by mutations in mtDNA. In some cases, the diseases, like mtDNA itself, are maternally inherited. In others, mutations appear to arise spontaneously during development, perhaps as the result of the action of a nuclear-gene product on the replication or recombination of the mtDNA molecule.

The biogenesis of a functional mitochondrion in all types of cells is a complex affair, with the information for most of the many hundreds of proteins required to construct the two membranes of the organelle coming from the nucleus (see ref. 9 for a recent review). Small though it is, human mtDNA makes an essential contribution to this process, as the figure indicates. The 16,596-base-pair molecule encodes the two structural RNAs of the mitochondrial ribosome, the 22 tRNAs required for mitochondrial protein synthesis and 13 proteins that, together with about 48 other proteins encoded by nuclear genes, make up the five multisubunit enzymes of the mitochondrial respiratory chain in the organelle's inner membrane (see also ref. 10).

One of the many unusual features of mtDNA in most organisms is its capacity for rapid change. In mammals, despite the extremely tight packing of information for gene products, the molecule fixes mutations 10–12 times faster than nuclear DNA[7]. The true rate of mutation is probably much higher, as many mitochondrial mutations occurring in the germ line may prevent formation of functional ova/spermatozoa, or the development of zygotes after fertilization. In somatic cells, the effects of potentially deleterious mutations may not be immediately obvious. Human cells normally contain thousands of mtDNA molecules and phenotypic expression will depend both on the extent of segregation of the mixed population of normal and mutant mtDNAs and on the cell's need for mitochondrial function. In this last respect, studies with respiratory inhibitors suggest that the tissues most likely to be affected are those of the brain and the central nervous system, followed by skeletal muscle, heart, kidney and liver[7].

In contrast to the ease with which mitochondrial dysfunction can be detected in lower eukaryotes such as yeast and *Neurospora*, diagnosis of mitochondrial diseases in man is far from straightforward. A mitochondrial myopathy is suspected if patients display impaired muscle oxidative metabolism, detectable by exercise stress testing, or ^{31}P-NMR, but a full diagnosis generally depends on examination of biopsy material. Histological examination of Gomori trichrome-stained muscle tissue may reveal fibres with a mottled or irregular appearance and red staining of peripheral or intermyofibrillar zones (ragged red fibres). Electron microscopy often shows numerous, enlarged mitochondria, containing whorled inner membranes and crystalline inclusions of protein in the matrix or intracristal spaces. However, unambiguous proof of dysfunction usually requires isolation of mitochondria, followed by direct measurement of respiratory chain function.

In terms of clinical presentation, perhaps one of the simplest types of mitochondrial disease is Lebers hereditary



Simplified clinical map of human mtDNA, showing mutations associated with the mitochondrial myopathies described in the text. Genes encoded by the heavy (H) and light (L) strands are shown on the outer and inner of the two circles, respectively. tRNA genes are encircled and identified by their cognate amino acid, in the one-letter code. COI–III: genes for subunits I–III of cytochrome *c* oxidase. ATPase 6,8: genes for subunits 6 and 8 of the mitochondrial ATPase. ND1–6: genes for 6 subunits of the NADH dehydrogenase (complex I). Cytb: gene for cytochrome *b*. P_{H1} P_{H2} and P_L are promoters of transcription. O_H and O_L designate the origins of DNA replication for the two strands. LHON indicates the approximate position of the G to A substitution in the gene for ND4. Arcs represent the size and position of deletions found in various myopathies with stippled lines indicating regions of uncertainty. MM1, MM2 refer to the mitochondrial myopathies described in ref. 4; KSS to the Kearns–Sayres syndrome patients described in ref. 5; PEO to the progressive external ophthalmoplegia patients described in ref. 6; CPEO (M) and CPEO (D) to the mother and daughter with chronic progressive external ophthalmoplegia described in ref. 13.

optic neuropathy (LHON). The disease results in sudden loss of vision between adolescence and adulthood, due to death of the optic nerve. Affected relatives are invariably related through the maternal lineage. Using the polymerase chain reaction to facilitate analysis of segments of mtDNA from patients belonging to several different pedigrees, Wallace *et al.*[2] could show that each affected individual contains only a single type of mtDNA. Sequence analysis showed that a G to A transition, converting a highly conserved arginine residue to a histidine in the gene for subunit 4 of NADH dehydrogenase, was the only change that could be correlated with the presence of the disease in 33 individuals from 9 pedigrees. Although not yet formally proven, it seems reasonable to assume that this subtle missense mutation affects the efficiency of either electron flow or energy transduction through the enzyme, thus reducing ATP production enough within the optic nerve to cause premature cell death and blindness.

A more complex clinical picture is presented by patients with a maternally inherited mitochondrial encephalomyopathy that results in myoclonic epilepsy associated with ragged red fibres (MERRF). Patients display varying degrees of mitochondrial abnormality, as manifested by impaired muscle oxidative metabolism and reduced respiratory chain activity[3]. Generally speaking, the order of affected tissues is consistent with their reliance on mitochondrial function and the severity of the biochemical defects correlates with the severity of the symptoms. Molecular analysis of mtDNA, although excluding large deletions or insertions, has so far failed to pinpoint the change responsible. The wide range of clinical symptoms suggests, however, that the patients are in fact mosaics of wild type and mutant mtDNAs, the relative proportion of each type of molecule varying along the lineage, so determining the severity of the affliction. The disease is thus most probably caused by one or more deleterious mutations, that can only be maintained in the heteroplasmic state, in association with wild type mtDNA.

Heteroplasmy of mtDNA is a common feature of a number of other mitochondrial abnormalities associated with a wide range of neuromuscular diseases, including the Kearns–Sayres syndrome (KSS) and chronic progressive external ophthalmoplegia (CPEO). In these cases, patients' muscle tissue contains mtDNA molecules with large-scale deletions extending over several genes[4,5,8]. These mutant molecules are always mixed with normal mtDNAs and represent 20–90 per cent of the total number of molecules. Curiously, there is no obvious correlation between the severity of the clinical symptoms or biochemical abnormality and either the location of the deletion or the number of deleted genes. But as transcription of mtDNA generates polycistronic transcripts corresponding to the full length of each DNA strand, a major deletion anywhere in the genome may have serious consequences for expression of genes not covered by the deletion.

In contrast to LHON and MERRF, diseases associated with deleted mtDNAs have usually not been inherited. Usually, deleted molecules within the muscle fibres of an individual have the same structure, suggesting that they arise by the clonal amplification of a single spontaneous mutational event that probably took place at some stage during fetal development. The pleiotropic nature of these diseases, thus, is likely to depend on the relative proportion of normal and deleted mtDNAs in different organs. This view is supported by recent findings[11] that a deletion in mtDNA is the probable cause of a strikingly different human disease, namely Pearson's syndrome, a fatal disorder of the haemopoietic system.

Of interest for the mechanism of mutation is the fact that all deletions so far analysed in detail have their endpoints in short repeated sequences, which seem to be especially abundant in human mtDNA. One site in particular, consisting of direct repeats of 13 base pairs separated by approximately 5 kilobases, seems to be a hot spot for mutation. In a study of 30 patients with KSS or CPEO, 12 were found to have identical deletions with endpoints in a direct repeat of 13 base pairs[5], whereas an identical deletion is found in a patient suffering from Pearson's syndrome[11], suggesting that this site is a hotspot for mutation. Intramolecular homologous recombination or slipped mispairing during replication, which are both plausible mechanisms for generation of large deletions in other DNAs, may be a significant source of the deletions in these diseases, but neither explanation is really satisfactory.

First, there are no reports of any DNA-repair phenomena in animal mitochondria and orthodox recombination mechanisms appear to be absent[12]. Second, slippage mispairing, which depends on the generation of single-stranded regions in complementary strands carrying direct repeats, may not apply to animal mtDNAs, which are replicated from two origins, one on each strand, by a mechanism that is not expected to expose long stretches of complementary single-stranded DNA at any stage. But as Holt *et al.* have suggested[8], a discontinuous synthesis mechanism may operate during periods of rapid mtDNA replication, such as during oogenesis when many of the mutations may occur.

Also of interest in terms of mutational mechanisms is the finding[13] of a mother and daughter suffering from CPEO who possess mtDNA deletions of different sizes (2.5 and 5.0 kilobases). Thus, either the female germ line of this lineage contains two different deleted mtDNA molecules or, more likely in view of the complexity of interactions between nuclear and mitochondrial genomes during mitochondrial biogenesis, some other hereditary factor predisposes the mtDNAs of these patients to deletion.

Involvement of nuclear factors is also suggested by the results of a recent study[6] of an Italian family containing several individuals affected by a late-onset mitochondrial myopathy with symptoms similar to those of KSS. Like KSS and

DNA amplification

"The principles for extensive synthesis of the duplexed tRNA genes which emerge from the present work are the following. The DNA duplex would be denatured to form single strands. This denaturation step would be carried out in the presence of a sufficiently large excess of the two appropriate primers. Upon cooling, one would hope to obtain two structures, each containing the full length of the template strand appropriately complexed with the primer. DNA polymerase will be added to complete the process of repair replication. Two molecules of the original duplex should result. The whole cycle could be repeated, there being added every time a fresh dose of the enzyme. It is however, possible that upon cooling after denaturation of the DNA duplex, renaturation to form the original duplex would predominate over the template-primer complex formation. If this tendency could not be circumvented by adjusting the concentrations of the primers, clearly one would have to resort to the separation of the strands and then carry out repair replication. After every cycle of repair replication, the process of strand separation would have to be repeated. Experiments based on these lines of thought are in progress."

Members of the growing band of biomedical scientists who use PCR — the 'polymerase chain reaction' for amplifying specific DNA sequences, described by Kary Mullis and colleagues at the Cetus Corporation in 1985 — may like to dust off the *Journal of Molecular Biology* for 1971 (vol. 56, pages 341–361) and read a paper from H. Gobind Khorana and colleagues, the last paragraph of which is reprinted here. □

100 years ago

From *Nature* **40**, cci; 17 October 1889.

CPEO patients, these individuals are heteroplasmic, with the mutant mtDNA molecules containing large deletions whose boundaries lie in short repeated sequences. But, whereas KSS and CPEO arise sporadically, this particular disease is inherited. Moreover, it is inherited in an autosomal dominant fashion, suggesting that a mutation in a nuclear-coded protein may accelerate the accumulation of mitochondrial mutations. All deletions in these patients have their 5′-ends located close to the D-loop, a region of mtDNA that plays a crucial role in its replication and transcription, and hence is likely to be a site of important DNA–protein interactions. Strikingly, the deletions in different patients display different 3′-end-points, suggesting that the heteroplasmic mtDNA populations are not transmitted *per se*, but are generated *de novo* by a common mechanism.

What lessons can be learnt so far? The first is that the findings bring a unifying picture to a diverse group of human diseases. The concepts developed should simplify the diagnosis of even mildly affected individuals and they will probably allow the recognition of new forms of disease, whose origins lie in mitochondrial dysfunction. Of particular importance in this respect are late-onset degenerative diseases and possibly also those associated with ageing[14]. The high rate of mutation means that mtDNA will accumulate mutations during the whole life of an individual, thus giving rise to a range of cells with varying capacity for energy production and survival. The recognition of the mitochondrion as a common factor in such diseases should, however, also facilitate efforts at therapy. As initial promising results indicate (reviewed in ref. 10), it should be possible to counteract the effects of the mutations by stimulating cellular energy production by administration of cofactors and redox compounds. Finally, further study of those diseases in which gene expression in mitochondria is affected by changes in nuclear-encoded proteins should yield insight into the nature of the cross-talk between the two genetic systems that controls the biogenesis of this fascinating organelle. □

L.A. Grivell is in the Department of Molecular Cell Biology, University of Amsterdam, Amsterdam 1098 SM, The Netherlands.

1. Borst, P. & Grivell, L.A. *Nature* **290**, 443–444 (1981).
2. Wallace, D.C. *et al. Science* **242**, 1427–1430 (1988).
3. Wallace, D.C. *et al. Cell* **55**, 601–610 (1988).
4. Holt, I.J. *et al. Nature* **331**, 717–719 (1988).
5. Schon, E.A. *et al. Science* **244**, 346–349 (1989).
6. Zeviani, M. *et al. Nature* **339**, 309–311 (1989).
7. Wallace, D.C. *Trends Genet.* **5**, 9–13 (1989).
8. Holt, I.J., Harding, A.E. & Morgan-Hughes, J.A. *Nucl. Acids Res.* **17**, 4465–4469 (1989).
9. Grivell, L.A. *Eur. J. Biochem.* **182**, 477–493. (1989).
10. Capaldi, R.A. *Trends Biochem. Sci.* **13**, 144 (1988).
11. Rotig, A. *et al. Lancet* **I**, 1902–1903 (1989).
12. Clayton, D.A. *Cell* **28**, 693–705 (1982).
13. Ozawa, T. *et al. Biochem. biophys. res. Commun.* **154**, 1240–1247 (1988).
14. Linnane, A.W. *et al. Lancet* **I**, 642–645 (1989).

THEORETICAL PHYSICS

Is quantum mechanics linear ?

Richard Thompson

QUANTUM mechanics has been subjected over the years to many stringent tests, in which theoretical predictions for specific systems have been compared with experimental measurements to high precision. Really, however, these should be seen as tests of the accuracy of the hamiltonian equation, which describes a particular physical system, rather than of quantum mechanics itself. Earlier this year, Weinberg[1] suggested a way of testing whether nonlinear corrections to the framework of quantum mechanics are necessary. Applying his method to data from an ion-trap atomic-clock experiment at the National Institute of Standards and Technology (NIST) in Boulder, Colorado, he found a minuscule upper limit for any such corrections. The NIST researchers now report[2] a new experiment, based on Weinberg's approach, designed to reveal any nonlinearities; their upper limit is a further five orders of magnitude smaller than Weinberg's.

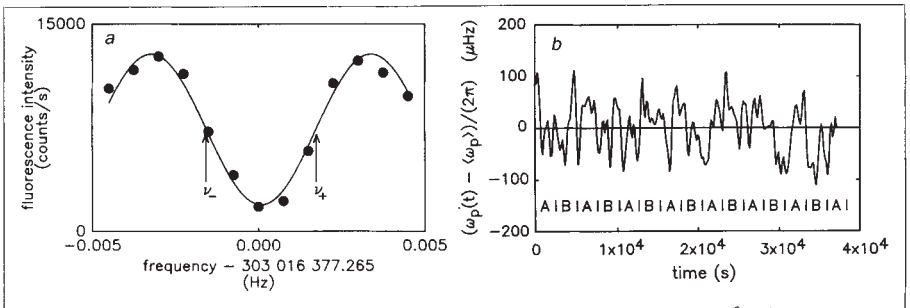
The existence of nonlinearities in quantum mechanics could have profound consequences for theoretical physics: the superposition principle, for example, could be vulnerable. Certainly, it is important that the very formalism of quantum mechanics should be tested.

The nonlinearity now being investigated is similar to that seen in all nonlinear oscillators for which the resonant frequency depends on the amplitude of the oscillations excited. A pendulum provides an example of such a system: as Galileo observed, for small oscillations the period is constant; but for large angular amplitudes, the period becomes longer. For a simple two-level quantum-mechanical system, the nonlinearity results in a change of the resonant frequency as the degree of excitation into the upper level is changed. A two-level system is equivalent to a spin one-half particle in an external magnetic field, for which the degree of excitation is related to the 'tipping angle' θ, which is the angle between the magnetic field and the spin. Here, $\theta=0$ corresponds to the pure ground state and $\theta=\pi$ to complete excitation into the upper state. Thus in this case, the precession frequency would not be constant, as predicted in linear quantum mechanics, but would be a function of θ.

This nonlinearity should not be confused with that of, for example, a material which exhibits nonlinear optical phenomena at high light intensities. That system and its nonlinear response to light is described by (linear) quantum mechanics, depends on the material's properties and is well understood. Nonlinear quantum mechanics would result in transition frequencies that are a function of the degree of excitation of the system, showing some similarity with the predictions of earlier neoclassical theory reviewed by Mandel[3].

In his original paper[1], Weinberg showed that one weak driving frequency would not be able to drive the whole system from an initial (pure) state to the other state, because the system would become detuned during the process, owing to the nonlinearity. Only if this detuning were less than $1/t$, where t is the time taken for the excitation, would complete excitation be possible. Thus, if the complete transition is observed, this sets a limit on the nonlinearity present.

To understand Weinberg's approach, consider first the form of the standard Schrödinger equation used in quantum mechanics: $i\hbar d\psi/dt=H\psi$. The idea is to find a wavefunction ψ, which is a (possibly complicated) function of space and time, that satisfies the equation. H, the hamiltonian, depends on the interactions in the system — for example, the Coulomb attraction between the nucleus and electron in an atom — and the equation is linear in ψ. ▶



Search for nonlinearities. *a*, Ramsey signal of the 'clock' transition of $^9Be^+$. When the microwave pulses of the Ramsey probe are tuned to the precession frequency ω_p, the maximum number of ions are transferred from the ion's fluorescing state, so that the detected laser-induced fluorescence is minimized. *b*, Be^+ precession frequency $\omega_p(t)$, referred to a passive hydrogen maser, as a function of time. The periods marked A (B) correspond to the tipping angles θ_A (θ_B). No systematic variation is discernible. (From ref. 2.)

Weinberg recasts this equation in such a way that H is a function of ψ and its complex conjugate ψ^*. The approaches are equivalent if H is linear in the product of ψ and ψ^*. With higher powers of ψ and ψ^* involved, the equation becomes nonlinear. Weinberg maintains certain conditions that apply to normal quantum mechanics in investigating the implications of the generalized equation: for example, the hamiltonian must be 'homogeneous' so that systems separated from one another are treated properly.

Although this formulation has been designed to add nonlinearity as generally as possible, it has been pointed out[4] that if the other postulates of quantum mechanics are left intact, it leads to violation of the second law of thermodynamics, which states that entropy always increases. Weinberg[5] replies that the definition of entropy used in ordinary quantum mechanics cannot be taken over into its nonlinear generalization, so a new definition must be sought. Thus, at present there does not appear to be any reason to exclude nonlinear corrections on theoretical grounds.

In general one would expect no more than a small nonlinear correction to the normal hamiltonian to arise — large corrections would have become apparent long ago. The most simple choice for the correction in the case of a spin-half particle in a magnetic field yields a precession frequency ω_p given by

$$\omega_p = \omega_0 - 4(\varepsilon/\hbar)\cos^2(\theta/2)$$

where θ is the tipping angle and ε measures the strength of the nonlinearity (ω_0 is the precession frequency normally predicted and $\hbar$ is Planck's constant).

The experimental system used for this investigation is a cloud of trapped and cooled beryllium ions[6], exposed to two coherent bursts of microwave radiation separated by a long period of free precession (the Ramsey technique; see the News pages for more on Ramsey's Nobel-prize winning work on atomic spectroscopy). The Penning ion trap, which uses a combination of static electric and magnetic fields to confine the ions to a small volume, contains of the order of 10,000 $^9Be^+$ ions. These are maintained at low temperatures (about 0.25 K) by collisions with roughly 100,000 $^{26}Mg^+$ ions held in the same trap. The ions are laser cooled (see my previous News and Views article[7]) by interaction with a laser beam tuned to a frequency just below their resonance line at 280 nm. Each time an ion absorbs a photon from the beam it slows down by an amount corresponding to the absorbed photon momentum until sub-kelvin temperatures are reached. The Be^+ ions are then cooled 'sympathetically' by the cold Mg^+ ions. The lifetime of the ions in the trap is essentially infinite as the whole system is maintained in an ultra-high vacuum (at pressures below 10^{-13} atmospheres).

The transition investigated is a nuclear spin flip in the Be^+ ions, which is field independent to first order at B=0.82 tesla and has a frequency of roughly 303 MHz. It is probed by irradiating the ion cloud with two coherent pulses (which would usually be $\pi/2$ pulses, changing θ by $\pi/2$ radians) separated by a period of the order of 100 seconds. It is only in an ion-trap experiment that such long interaction times and low temperatures can be achieved and this is what makes the measurement so sensitive. After irradiating the ions with the 303 MHz radiation, the number of ions to have made the transition is determined by measuring the population left in the initial state from the amount of fluorescence produced when the ions are irradiated by light at the Be^+ resonance wavelength of 313 nm.

The point of the experiment is to determine a difference in the precession frequency for different values of θ, so the first coherent pulse used in the Ramsey probe is chosen to give either $\theta=\theta_A=1.02$ radians or $\theta=\theta_B=2.12$ radians. Using a conventional feedback system, the probe oscillator is locked to the ion's precession frequency with $\theta=\theta_A$ for roughly 40 minutes, then with $\theta=\theta_B$ for 40 minutes and so on. A sample plot of this locked frequency against time is shown in the figure and does not show any clear oscillation between two values as θ is changed, as would be expected if the precession frequency were changing with θ in the manner predicted.

The accuracy of the experiment is limited by the frequency reference available and the detected frequency shift, which is consistent with zero, is 3.8 ± 8.3 μHz (the error corresponds to less than one complete cycle per day), yielding a value for ε of 1.8 ± 4.0 μHz. Thus the maximum possible value of ε is equivalent to an energy of 2.4×10^{-20} eV, or less than 4 parts in 10^{27} of the binding energy per nucleon of the beryllium nucleus. This limit is 10^5 times smaller than the previous one and can still be improved by roughly an order of magnitude with the present apparatus. Therefore, at the moment there is no reason to suppose that the framework of quantum mechanics shows nonlinearity at a significant level, though doubtless the search for such effects will continue. □

Richard Thompson is in the Blackett Laboratory, Imperial College of Science, Technology and Medicine, Prince Consort Road, London SW7 2BZ, UK.

1. Weinberg, S. *Phys. Rev. Lett.* **62**, 485–488 (1989).
2. Bollinger, J.J. *et al. Phys. Rev. Lett.* **63**, 1031–1034 (1989).
3. Mandel, L. in *Progress in Optics* XIII (ed. Wolf, E.) 27–68 (North Holland, Amsterdam, 1976).
4. Peres, A. *Phys. Rev. Lett.* **63**, 1114 (1989).
5. Weinberg, S. *Phys. Rev. Lett.* **63**, 1115 (1989).
6. Bollinger, J.J. *et al. Phys. Rev. Lett.* **54**, 1000–1003 (1985).
7. Thompson, R. *Nature* **335**, 588–589 (1988).

DAEDALUS

Stirrings of life

LAST WEEK Daedalus invented his circularly polarized microwave reactor. It spins the polar molecules in a solution, driving them along like little propellers if they have the right asymmetry. The slow, diffusion-limited kinetics of chemical reactions is thus vastly accelerated. The chemistry of life, dominated by polar chiral molecules, should be particularly speeded up.

DREADCO's botanists are therefore constructing a novel 'Microtational forcing house' for plants. A powerful circularly polarized microwave field should speed their growth tremendously. Provided with extra warmth, intense light, concentrated nutrients and a carbon-rich atmosphere to fuel their furious growth, they should hustle from seed to maturity in a mere few hours or days. Indeed, when perfected, the new forcing house may be best laid out as a conveyor-belt production line, with seeds being continuously set on the belt at one end of the building and harvested as fully grown plants as they emerge at the other. Agriculture should be wonderfully compressed in space and time, releasing vast tracts of land for airports, motorways, leisure centres, theme parks and other progressive amenities.

The biochemical effects of circularly polarized microwaves on human and animal growth should be equally dramatic, and DREADCO's biologists are exploring them cautiously. Animal growth and metabolism should be directly accelerated. Microtational pig sties, cow sheds and hen houses should bring livestock to rapid maturity without the problems and side effects of hormone treatment. Medicine may gain as well. DREADCO's microtational sick bay could well speed up the healing of wounds, the regrowth of lost hair or nails and the defeat of infection. (On the other hand, bacterial multiplication may be accelerated even more.) Impatient body builders should also welcome a microtational gymnasium.

More revolutionary still, a microtational pregnancy smock might deliver an accelerated baby in far less than the traditional nine months. But a microtational cradle or nursery, to speed its subsequent physical progress, could have grave drawbacks. The long drawn out human childhood has evolved to give the growing child time to absorb the vast amount of information it will need as an adult. And learning, like all mental activities, depends ultimately on the speed of neural transmission — an ionic process unaffected by circularly polarized microwaves. So an accelerated child would gain rapidly in stature, but not in wisdom. He would reach physical adulthood while mentally still very childish and backward. He might also be biologically that much older and would grow old and die before his chronological time. David Jones

SIV adaption to human cells

SIR—Molecular cloning of simian immunodeficiency virus derived from rhesus macaques (SIV_{MAC}) has revealed[1-4] an in-frame stop codon in the envelope gene that truncates the cytoplasmic domain of the transmembrane glycoprotein relative to that of the human immunodeficiency virus HIV-1. As a result, SIV_{MAC} expresses a transmembrane glycoprotein of M_r 32,000 (gp32), rather than the 41K transmembrane glycoprotein of HIV-1 (gp41). We investigated the importance of the stop codon by removing it from an infectious molecular clone of SIV_{MAC} (pBK28; designated pSIV-tm32). Our findings indicate that this stop codon is an adaptation of a non-human primate lentivirus to growth in human cell lines.

We removed the stop codon from the *env* gene of pSIV-tm32 by replacing a 78-base-pair (bp) fragment with a synthetic 75-bp fragment that lacked the stop codon (see *a* in figure). This generated a clone designated pSIV-tm41. pSIV-tm32 and pSIV-tm41 were transfected in parallel into HUT-78 cells and equivalent amounts of reverse transcriptase activity were detected in culture supernatants 10 days later. Analyses of SIV-specific glycoproeins revealed the presence of an elongated *env* precursor (gp 160+) in cells transfected with pSIV-tm41, and a smaller gp160 in cells transfected with pSIV-tm32 (see *b* in figure). Transmembrane proteins of the predicted size were observed by western blot analysis.

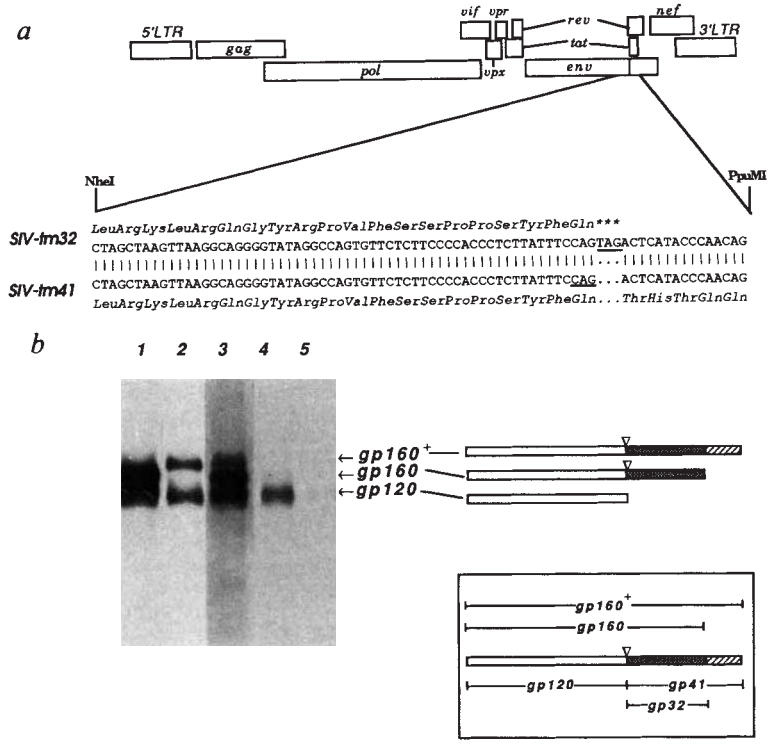
Filtered supernatants derived from transfected cells were used to infect HUT-78 cells. Infectivity titres of these stocks in HUT-78 cells were 10^4 $TCID_{50}$ ml^{-1} for SIV-tm32 and $<10^1$ $TCID_{50}$ ml^{-1} for SIV-tm41. Paradoxically, both stocks appeared to have roughly equivalent amounts of virus particles as judged by reverse transcriptase activity and gag protein production. A persistent infection was established in only one of eight attempts to infect HUT-78 cells with cell-free SIV-tm41. This culture (referred to as SIV-tm41/32) produced gp160+, and also gp160 (see figure), as expected for SIV-tm32. To investigate this unusual finding, we used polymerase chain reaction (PCR) to amplify the region surrounding the original SIV-tm32 stop codon and determine the DNA sequence (see table). This analysis revealed a C → T point mutation in the CAG codon immediately preceeding the original stop codon, thereby regenerating a stop codon in some viral genomes in this culture. Selection of a stop codon mutant indicated that tm32 conferred a growth advantage to SIV in the human cell line, HUT 78.

The host range of SIV-tm32 and SIV-tm41 was evaluated in additional human cells (H9 and peripheral blood mononuclear cells (PBMC)) and rhesus macaque PBMC. SIV-tm41 grew preferentially in rhesus PBMC. By contrast, SIV-tm32 was stably propagated in human cells. PCR analysis of SIV-tm32-infected rhesus PBMC revealed mutants lacking the stop codon by day 17 (see table). Thus, forms of the SIV transmembrane glycoprotein are selected based on the host cells used for virus propagation; an extended glycoprotein is preferred in rhesus PBMC and a truncated one in human cells.

We next sought to determine which transmembrane glycoprotein structure(s) are present in SIV-infected macaques. Four rhesus macaques were inoculated with SIV-tm32 or the SIV-tm41/32 virus mixture described above. All animals became persistently infected. Genomic DNAs from PBMC collected 10 months post-inoculation were analysed by PCR for stop codons (see table). Regardless of the original inoculum, the dominant genotype detected in circulating PBMC was SIV-tm41. By contrast, co-cultivation of human cells with PBMC from one infected macaque resulted in re-selection of the SIV-tm32 genotype. Thus, tm41 is the naturally occurring form of the SIV_{MAC} transmembrane glycoprotein in infected macaques. This conclusion is consistent with the observation that SIV-infected macaques have antibodies reactive to synthetic peptides derived from the region C-terminal to the stop codon of the transmembrane protein[5].

We have observed a similar conversion from a 41K to a 32K transmembrane glycoprotein in human cells (H9) transfected with a molecular clone of SIV from sooty mangabeys (SIV_{SM}), an African SIV closely related to SIV_{MAC}[6]. This conversion was due to an out-of-frame duplication (130 bp) in the region of the stop-codon in SIV_{MAC}. Truncation of the transmembrane glycoprotein has also been observed in SIV from African green monkeys SIV_{AGM}[7], and in HIV-2[8,9]. The form of transmembrane glycoprotein genes in SIV_{AGM} or HIV-2 infected individuals has not been determined. Similarly located stop codons have not been observed in HIV-1 clones. Furthermore, deliberate truncation of the HIV-1 glycoprotein in an analogous position seems to destroy its infectivity for HUT-78 cells (data not shown).

The cytoplasmic domain of viral envelope proteins has been implicated in virus assembly, cytopathic effects, and *env* protein transport to the cell surface[10-14]. We have not identified the biochemical basis for the selective growth characteristics of SIV with differing length *env* cytoplasmic domains. However, apparent levels of SIV-tm41 *env* glycoproteins in



a, Construction of SIV-tm 41. A map of open reading frames of the SIV genome is shown. The nucleotide and deduced amino-acid sequences of the synthetic restriction fragment used to remove the SIV-tm32 stop codon is shown below, aligned with the sequence of the original clone pSIV-tm32. pSIV-tm41 was constructed by ligation of a synthetic oligonucleotide lacking the stop codon into pSIV-tm32. *b*, Analysis of SIV glycoproteins. The env proteins of SIV-tm32 (lane 1), SIV-tm41 (lane 2), and the SIV-tm41/tm32 revertant virus mixture (lane 3) are shown following lentil lectin affinity purification and radioimmunoprecipitation[18]. Extracellular glycoprotein (gp120) of SIV-tm32 (lane 4) and SIV-tm41 (lane 5, faint band on this exposure) from purified virions are indicated. A diagram indicating the derivation of these proteins is also shown.

human cells are reduced relative to other viral proteins (data not shown), indicating that impaired replication of this virus in human cells may be the result of decreased stable *env* glycoprotein expression.

Our findings demonstrate that SIV_{MAC} genomes encoding a truncated transmembrane protein are selected by their ability to grow in human cells in culture. Thus, *in vitro* propagation may select for mutant forms of virus, skew the representation of genotypes within a virus mixture, and obscure the full potential of viral genomes by altering open reading frames. Tissue culture adaptation could conceivably select against AIDS pathogens (both SIV and HIV) by fortuitious host cell selection, defective replication, or differential cytopathic effects. For example, feline and murine leukaemia retroviruses which causes AIDS-like diseases in cats and mice, respectively, are replication defective. However, their closely related but minimially pathogenic helper viruses are replication competent[15–17]. In addition, less cytopathic viruses may predominate in surviving cells as highly cytopathic variants kill their host cell more efficiently. This report highlights the problems inherent in studying complicated host–pathogen relationships in tissue culture systems. The SIV-tm32/41 adaptation we described here does not appear to effect pathogenicity. However, molecular clones of SIV derived from tissue-culture-adapted virus are minimally pathogenic in macaques. Thus, functional analysis of viral genomes derived directly from tissues will be necessary to explore fully the pathogenic spectrum of AIDS viruses.

V. M. HIRSCH*†
P. EDMONDSON*‡
M. MURPHEY-CORB§
B. ARBEILLE¶
P. R. JOHNSON*
J. I. MULLINS*‡

*Department of Cancer Biology, Harvard School of Public Health, 665 Huntingdon Avenue, Boston, Massachusetts 021115, USA

†Department of Immunology, Georgetown University, NIH/Twinbrook II, 12441 Parklawn Drive, Rockville, Maryland 20852, USA

‡Department of Microbiology & Immunology, Stanford University School of Medicine, Palo Alto, California 94305-5402, USA

§Delta Regional Primate Research Center, Tulane University, Covington, Louisiana 70433, USA

¶Unité de Microscopie Électronique, UER Médecine, BP 3223, 37032 Tours, France

Genotypic changes in the SIV *env* gene following propagation in cell culture and *in vivo*

Infection of cultured cells Virus	Cell culture	Genotypes identified	Sequence surrounding the original stop codon
SIV-tm32			..TAT TTC CAG TAG ACT CAT ACC CAA CAG G..
	HuT78*	100% tm32 (4)	..TAT TTC CAG TAG ACT CAT ACC CAA CAG G..
	H9	100% tm32 (9)	..TAT TTC CAG TAG ACT CAT ACC CAA CAG G..
	HuPBMC	100% tm32 (20)	..TAT TTC CAG TAG ACT CAT ACC CAA CAG G..
	RhPBMC	27% tm32 (3)	..TAT TTC CAG TAG ACT CAT ACC CAA CAG G..
		73% tm41 (8)	..TAT TTC CAG TAG ACT CAT ACC CAA CAG G..
SIV-tm41			..TAT TTC CAG ACT CAT ACC CAA CAG G..
	HuT78†	76% tm41 (19)	..TAT TTC CAG ACT CAT ACC CAA CAG G..
		24% tm32 (6)	..TAT TTC CAG ACT CAT ACC CAA CAG G..
	H9	100% tm41 (20)	..TAT TTC CAG ACT CAT ACC CAA CAG G..
	HuPBMC	100% tm41 (20)	..TAT TTC CAG ACT CAT ACC CAA CAG G..
	RhPBMC	100% tm41 (20)	..TAT TTC CAG ACT CAT ACC CAA CAG G..
In vivo passage for 10 months in rhesus macaques			
Inoculum	Rhesus		
SIV-tm32	G014	19% tm32 (4)	..TAT TTC CAG TAG ACT CAT ACC CAA CAG G..
(HuT78)		81% tm41 (17)	..TAT TTC CAG CAG ACT CAT ACC CAA CAG G..
SIV-tm41/32	G442‡	81% tm41 (1)	..TAT TTC CAG ACT CAT ACC CAA CAG G..
(HuT78)		(20)	..TAT — — — — ACC CAA CAG G..
		19% tm32 (5)	..TAT TTC — — — — TAA CAG G..
	G594	100% tm41 (26)	..TAT TTC CAG ACT CAT ACC CAA CAG G..
In vitro passage in human cells of virus derived from a rhesus macaque			
Rhesus G442		100% tm32 (13)	..TAT TTC — — — — TAA CAG G..

Viral DNA sequences surrounding the original stop codon were amplified by PCR from infected cells in culture or uncultured PBMC, cloned and their DNA sequence determined. The percentage of the genotypes defined by the presence or absence of a stop codon (underlined) and the number of clones analysed (in parentheses) are indicated. Viral DNA sequences were amplifed from 1 μg cellular DNA during 35 heating–cooling–extension cycles (94 °C for 45s, 55 °C for 1 min, 72 °C for 1 min) in a Cetus-Perkin Elmer thermal cycler. Each 0.1-ml reaction contained 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 0.01% gelatin, 200 μM each of all four deoxynucleotide trisphosphates, 2mM $MgCl_2$, 2.5 U *Thermus aquaticus* polymerase (Cetus-Perkin Elmer), and 10 pmol of oligonucleotide primers. Samples (10 μl) from PBMC were reamplified in fresh reaction buffer for an additional 35 rounds with an internal, or 'nested' set of primers containing restriction sites at their 5° ends to facilitate later cloning. Products were cloned into either m13mp18 or pUC18 and sequenced using T7 DNA polymerase.
*Virus used to infect rhesus G014. †Virus used to infect rhesus G442 and G594. ‡Virus used to infect human cells following 10 months' propagation in rhesus G442.

1. Hirsch, V., Riedel, N., & Mullins, J.I. *Cell* **49**, 307 (1987).
2. Franchini, G. *et al. Nature* **328**, 539–543 (1987).
3. Chakrabarti, L. *et al. Nature* **328**, 543–547 (1987).
4. Kornfeld, H., Riedel, N., Viglianti, G.A., Hirsch, V., Mullins, J.I. *Nature* **326**, 610–613 (1987).
5. Franchini, G., Rushdie, J., O'Keefe, T. & Wong-Staal, F. *AIDS Res. Hum. Retrovir.* **4**, 243–259 (1989).
6. Hirsch, V., Olmsted, R., Murphey-Corb, M., Purcell, R., & Johnson, P.J. *Nature* **339**, 389–391 (1989).
7. Fukasawa, M. *et al. Nature* **333**, 457–461 (1988).
8. Guyader, M. *et al. Nature* **326**, 662–669 (1987).
9. Zagury, D. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
10. Perez, L.G. *et al. J. Virol.* **61**, 2981–2988 (1987).
11. Fisher, A.G. *et al. Science* **233**, 655–659 (1986).
12. Guan, J-L., Ruusala, A., Cao, H. & Rose, J.K. *Molec. Cell Biol.* **8**, 2869–2974 (1989).
13. Doyle, C., Sambrook, J., & Gething, M-J. *J. Cell Biol.* **100**, 704–714 (1986).
14. Sodroski, J. *et al. Science* **231**, 1549–1553 (1986).
15. Overbaugh, J., Donahue, P. Quackenbush, S.L., Hoover, E.A. & Mullins, J.I. *Science* **239**, 906–910 (1988).
16. Chattapadhyay, S.K., Morse, H.C., Makino, M., Ruscetti, S.K. & Hartley, J.W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3862–3866 (1989).
17. Aziz, D.C., Hanna, Z., & Jolicoeur, P. *Nature* **338**, 505–508 (1989).
18. Allan, J.S. *et al. Science* **228**, 1091-1094 (1985).

Too many leucine zippers?

SIR—The leucine zipper was first proposed[1] as a hypothetical structure facilitating dimer formation of the nuclear proteins C/EBP, Fos, Myc, Jun and GCN4. Its characteristic motif is a periodic repeat of leucines at every seventh position (a heptad repeat) in a segment of 22–29 residues, and its overall composition (in particular, the absence of prolines and glycines, and a high density of ion pairs) is compatible with α-helical secondary structure, suggesting that dimerization is based upon coiled coil interactions[1,2]. However, a distinction has been made between the leucine zipper and the coiled coil interfaces in fibrillar proteins like tropomyosin, keratin and lamins on the basis of the almost exclusive use of leucines in the $i+7$th positions and the non-conservation of hydrophobic residues in the $i+4$th positions (in Fos and Myc)[3]. More recently, it has been claimed that many other proteins contain the leucine zipper motif, including the fusion glycoproteins of paramyxoviruses[4], voltage- and ligand-gated ion channels[5] and glucose-transporter glycoproteins[6]. Note, however, that unlike the situation in C/EBP, the motif in the paramyxoviruses entails predominantly charged and uncharged polar residues in the $i+4$th positions, and the motif in the glucose-transporter proteins entails two

substitutions of isoleucine for leucine.

We wish to caution against the too ready acceptance of the significance of leucine heptad repeats found in databank searches. In particular, we address the following questions concerned with the statistical background of such searches. How many leucine heptad repeats might one expect to find merely by virtue of chance? Are leucine heptad repeats relatively more frequent in proteins than heptad repeats of other amino acids (taking into account the biases in residue composition)? Are heptad repeats of leucine more frequent than repeats of other period length?

In the table we list the probability of occurrence of 4 or 5 leucine repeats in random protein sequences of given size and leucine content. Natural protein sequences are not random, but the tabulated probability values may serve as a reference in estimating the statistical significance of the observed repeats. It is seen, for example, that 4 repeats in a 300–500 residue sequence of average leucine content (10%) occur with only 3–4% probability, but that this probability is considerably increased for sequences of higher leucine content or greater length.

We have screened in excess of 450 distinct mammalian proteins of mean length 450 residues and average leucine content 9.7% for occurrences of the motif L-X_6-L-X_6-L-X_6-L, where L is leucine and X is any other amino acid. The motif is found in more than 30 of the sequences, but in only about half of them, including ribophorin I, spectrin and interleukin-3, is X never a proline. Heptad repeats of other amino acids occur much less than half as frequently as the leucine heptads, in agreement with the fact that leucine is by far the most common amino acid over all the proteins analysed. The second most frequent heptad repeat (occurring more than five times as often as expected) contains glutamic acid residues instead of leucine (as, for example, in the $i+3$th positions of the Fos leucine zipper).

Probability of observing a success run of r periodically repeated leucines in a sequence of length n for a given frequency f of leucine in the sequence.

	$r = 4$			$r = 5$		
n/f	6.5%	10.0%	13.5%	6.5%	10.0%	13.5%
200	0.003	0.02	0.06	0.001	0.002	0.008
300	0.005	0.03	0.08	0.001	0.003	0.01
500	0.008	0.04	0.13	0.001	0.004	0.02
1000	0.02	0.09	0.25	0.001	0.009	0.04

The probability is closely approximated by $(1 - fx)/[(r + 1 - rx)(l-f)x^{n+1}]$, where x solves $(1 - f)x(1 + fx + .. + f^{r-1}x^{r-1}) = 1$ (ref. 7). A lower bound for the probability of observing an r-repeat of any (not predetermined) amino acid is given by the same formula with values $r - 1$ and $f = 5\%$ (0.03–0.06 for $r = 4$ and $n = 300$–500).

For the same set of proteins, leucine repeats of length 4 with spacing 5 or 7 (rather than 6) are found in about 20 proteins each. Also, spacings 3 and 6 are preferred over spacings 4,5,7 and 8 between (not necessarily nearest) neighbouring leucines. This relative excess of leucine heptad repeats is consistent with the role of leucines in establishing hydrophobic faces of α-helices, including those involved in coiled coil interactions. Interestingly, the same preference for spacing 6 holds for glutamic acid residues, possibly reflecting their role in establishing hydrophilic faces of α-helices.

Thus the leucine repeat by itself occurs widely on both probabilistic and empirical grounds. This abundance undoubtedly reflects the particular structural role of the motif but it also warrants a cautioning note with respect to the prolific citation of leucine zippers, particularly when pattern searches allow for substitutions in the leucine positions. While one might rightly want to accommodate true variations in primary sequence yielding equivalent three-dimensional structures, attention ought to be given to the concomitant increase in the level of false positive occurrences. Based on our probability estimates and data work, we think that perhaps as many as two of every three leucine zipper motifs are simply chance occurrences.

VOLKER BRENDEL
SAMUEL KARLIN
Department of Mathematics, Stanford University, Stanford, California 94305, USA

1. Landschulz, W.H., Johnson, P.F. & McKnight, S.L. *Science* **240**, 1759–1764 (1988).
2. O'Shea, E.K., Rutkowski, R. & Kim, P.S. *Science* **243**, 538–542 (1989).
3. Landschulz, W.H., Johnson, P.F. & McKnight, S.L. *Science* **243**, 1681–1688 (1989).
4. Buckland, R. & Wild, F. *Nature* **338**, 547 (1989).
5. McCormack, K *et al. Nature* **340**, 103 (1989).
6. White, M.K. & Weber, M.J. *Nature* **340**, 103–104 (1989).
7. Feller, W. *An Introduction to Probability Theory and its Applications* 3rd edn Vol. 1 325 (Wiley, New York, 1968).

Moving proteins

SIR—The use of Con A labelled with 40-nm gold particles to show forward movement of membrane glycoproteins in locomotor cells[1] is exciting, but it should not be forgotten that during the past decade, several groups in the leukocyte chemotaxis field have reported forward distribution of cell-surface receptors in polarized and actively moving leukocytes, albeit using less refined techniques. Many workers have considered that this distribution is the result of an anterior movement of membrane proteins rather than their reinsertion at the leading edge.

Anterior distribution has been reported for Fc receptors[2-4] and for receptors for a chemoattractant peptide[5]. A variety of functionally unrelated proteins, including Thy-1 (ref. 4) and CD45 (ref. 6) also show forward distribution on locomotor lymphocytes. This anterior distribution is independent of the ligand used to stimulate polarization and locomotion, and occurs rapidly[7]. We have suggested[8] that forward movement of chemotaxis receptors is an intrinsic component of chemoattractant-induced polarization, that it is intrinsic to the directional locomotor response, reinforcing the persistence of that response in cells moving up a gradient, and that it also accounts for the long stretches of persistent locomotion in a given direction seen in cells moving randomly when the chemoattractant is uniformly distributed.

A question remains about mechanism. Kucik *et al.* discuss myosin I which is linked to both actin and to membrane sites, as a possible motor. However, Thy-1, a protein which is inserted by a phosphatidyl-linkage into the outer leaflet of the lipid bilayer, nevertheless moves forward and maintains an anterior distribution in motile lymphocytes, indicating that linkage of cytoskeleton to forward-moving proteins may be indirect.

P.C. WILKINSON
Department of Bacteriology & Immunology, University of Glasgow, Glasgow G12 8QQ, UK
W.S. HASTON
Department of Microbiology, University of Aberdeen, Aberdeen AB9 2ZD, UK

1. Kucik, D., Elsen, E. & Sheetz, M. *Nature* **340**, 315 (1989).
2. Wilkinson, P.C., Michl, J. & Silverstein, S.C. *Cell Biol. Int. Rep.* **4**, 736 (1980).
3. Walter, R., Berlin, R. & Oliver, J. *Nature* **286**, 724 (1980).
4. Shields, J.M. & Haston, W.S. *J. Cell Sci.* **74**, 75–93 (1985).
5. Sullivan, S.J., Daukas, G. & Zigmond, S.H. *J. Cell Biol.* **99**, 1461–1467 (1984).
6. Haston, W.S. & Wilkinson, P.C. *Curr. Opinion Immunol.* **1**, 5–9 (1988).
7. Haston, W.S. & Shields, J.M. *J. Cell Sci.* **82**, 249 (1986).
8. Haston, W.S. & Wilkinson, P.C. *J. Cell Sci.* **87**, 373 (1987).

SIR—Kucik *et al.*[1] state that "no forward transport of particles has previously been reported" on locomoting cells and, again, that "only rearward transport of surface particles had been observed". This is not the case. As detailed in my recent review[2] of the entire literature on the use of inert particles as markers for cell surface membrane dynamics during cell motility, workers have observed the forward movement of inert marker particles (carmine, India ink, carbon particles) along the surface of locomoting freshwater amoebae.

R. A. BLOODGOOD
Department of Anatomy and Cell Biology, University of Virginia School of Medicine, Charlottesville, Virginia 22908, USA

1. Kucik, D., Elsen, E. & Sheetz, M. *Nature* **340**, 315 (1989).
2. Bloodgood, R. A. in *Microspheres: Medical and Biological Applications* (eds Rembaum, A. & Tokes, Z.) 165–192 (CRC, Boca Raton, 1988).

Mass loss in PSR 1957+20

SIR—Although the precise mechanism of the eclipses in the binary system containing pulsar PSR 1957+20 is still unclear[1-5], it is generally agreed[1-4] that the $\sim 0.02M_{\odot}$ companion[1] in circular orbit ($e < 0.0002$) (ref. 2) is losing mass. Measurement of the mass-loss rate of this very light companion can provide a critical test of theories of this millisecond pulsar as well as the theory of millisecond pulsar evolution[3,4,6,7]. As pointed out by Kluźniak *et al.*[3], the mass loss need not lead to a significant change of the orbital period, P_{orb}, if the ratio of the specific angular momentum lost to the specific angular momentum of the companion star, k, is close to unity, because

$$\dot{P}_{orb}/P_{orb} = 2(k-1)\dot{m}_c/m_c \qquad (1)$$

where m_c is the mass of the companion. Here we point out that if the change in speed of the neutron star, v_1, another observable quantity, were measured, then the mass-loss rate could be deduced as

$$\frac{\dot{m}_c}{m_c} = \frac{\dot{v}_1}{v_1} + \frac{1}{3}\frac{\dot{P}_{orb}}{P_{orb}} = \frac{\dot{a}_1}{a_1} - \frac{2}{3}\frac{\dot{P}_{orb}}{P_{orb}} \qquad (2)$$

where a is the semimajor axis.

Equation (2) is model independent. Therefore, observational determination of mass-loss rate can be made by measuring v_1 and P_{orb} as well as their time derivatives.

From theoretical arguments, if the evaporation of the companion is fast enough to explain the original solitary millisecond pulsar PSR 1937+214, $m_c/\dot{m}_c < 10^8$ yr is needed. Six months of data on PSR 1957+20 gives $3P_{orb}/\dot{P}_{orb} > 10^7$ yr (ref. 2), while $v_1/\dot{v}_1$ is easily measured from the Doppler shift of the pulse period. Within five years, we would probably reach the limit of theoretical interest even if $\dot{P}_{orb}/P_{orb}$ is the dominant term.

KAIYOU CHEN

Institute for Theoretical Physics,
University of California,
Santa Barbara,
California 93106, USA

1. Fruchter, A.S., Stinebring, D.R. & Taylor, J.H. *Nature* **333**, 227 (1988).
2. Fruchter, A.S. *thesis* (Princeton Univ, 1989).
3. Kluźniak, W., Ruderman, M., Shaham, J. & Tavani, M. *Nature* **334**, 225 (1988).
4. Phinney, E.S., Evans, C.R., Blandford, R.D. & Kulkarni, S.R. *Nature* **333**, 832 (1988).
5. Michel, F.C. *Nature* **337**, 236 (1988).
6. Ruderman, M., Shaham, J. & Tavani, M. *Astrophys. J.* **336**, 507 (1989).
7. Kulkarni, S.R. & Narayan, R. *Astrophys. J.* **335**, 755 (1988).

Kinematics of the Alpine arc

SIR—Platt *et al.*[1] provide a compilation of kinematic data for the Alpine arc which will find far wider use than their particular purpose of establishing the independence of the Adria plate from both the European and African plates. They stress that "we have included only data for which we could find good evidence of age". Their ages are out of date, however, in that they have not taken account of some recent research or of changes in stratigraphic correlations. This is most clearly the case in the south-west Alps, where they have allocated all their south-west-directed motions in the external zones and in the nappes of Embrunais–Ubaye (their arrows 37,40,45,46,47,49–51) to ages that are post-Eocene.

A principally sedimentological study[2] in an area between the Embrunais–Ubaye nappes and the Provençale platform found south-west-directed thrusts that had cut the surface during deposition of calcaire nummulitique (Eocene). The study shows that south-west-directed thrusting constrains the geometry of the Grès d'Annot basin (Eocene). Most thrusts in the upper parts of the external zone succession, and many in the sub-Briançonnais of the lower Embrunais–Ubaye nappes[3] are folded by structures that formed during nappe emplacement[3,4]. The hypothesis that the nappe loading caused the thrusts in the external zone beneath and which has generally been accepted as a basis for structural correlation, is therefore not valid and an earlier (Eocene) age for much of the thrusting is established.

It is also necessary to reinterpret ages in the literature. Many biostratigraphic zones were once considered of Oligocene age, but are now firmly established as Eocene (see ref. 5, for example). Whereas the shift of the zonal scheme has been 'absorbed' relativly easily, many early attributions of Oligocene age to structural events persist in the structural geology literature. The emplacement of the lower Embrunais–Ubaye nappes is important in structural correlation and is well constrained stratigraphically.

Referring to the nannoplanktonic (NP) zonal scheme, the topmost beds of the Grès d'Annot are NP 19–20 (Muller, cited in ref. 5). The nappes override these beds, after deposition of an olistostrome[6], but were uplifted, subjected to subaerial erosion, and become the source of the sediment near the base of zone NP 22 at Barrême[7]. Emplacement was therefore within a time window corresponding to zones NP 20 and NP 21. If we now accept the zonal boundary NP 21–22 as the Eocene/Oligocene boundary (as for example p. 393 of ref. 5), the nappe emplacement and all associated structures must be of Eocene, not Oligocene, age.

The pitfalls of structural correlation in the Embrunais–Ubaye region are discussed elsewhere[8]. Readers of Platt *et al.*[1] are advised that these amended ages do not invalidate the general time-divisions of their Figs 1 and 2 but that, in the south-west Alps and possibly elsewhere, some of the data which they have treated as post-Eocene may be of Eocene age. In any event, their data set is misleading in that it does not indicate the scale and complexity of Eocene Alpine tectonics which, in the south-west Alps at least, was dominated by southwestward displacements[8].

NORMAN FRY

Department of Geology,
University of Wales College of Cardiff,
PO Box 914,
Cardiff CF1 3YE, UK

PLATT REPLIES—The evidence cited by Fry (unpublished when we wrote our paper[1]) for a late Eocene onset of deformation in the external zone of the south-west Alps is of considerable interest, as precise timing of the beginning of deformation in different parts of the Alps should help constrain the rate of convergence. The main evolution of the thrust belt in this region, however, is clearly post-Eocene: Grès d'Annot and flysch of Priabonian (late Eocene) age are imbricated in the thrust sheets; and in the Digne and Nice arcs, Triassic rocks have been thrust over sediments of late Miocene to Pliocene age.

The presence of detritus derived from the Embrunais–Ubaye nappes in late Eocene sediments is not a valid constraint on the timing of external-zone deformation. These nappes, which form part of the internal zone, had been incorporated into the Alpine orogenic wedge and could have been subject to erosion by late Eocene time: kinematic data reflecting this deformation are shown on our map (data sets 41, 42 and 44 in Fig. 1 of ref. 1) and generally trend west or north-west. The accretion of external-zone rocks beneath these nappes occurred later and caused folding, reimbrication and back-thrusting in the older parts of the wedge (data sets 43, 45 and 46): these structures are generally associated with north-east–south-west-directed kinematic indicators, parallel to those associated with the co-eval south-west-directed thrusting in the external zones (data sets 37, 47 and 49).

J. P. PLATT

Department of Earth Sciences,
Parks Road,
Oxford OX1 3PR, UK

1. Platt, J.P., Behrmann, J. H., Cunningham, P. C., Dewey, J. F., Helman, M., Parish, M., Shepley, M. G., Wallis, S. & Weston, P. J. *Nature* **337**, 158–161 (1989).
2. Apps, G. thesis, Univ. Liverpool (1987).
3. Lawson, K.D. thesis, Univ. Coll. Swansea (1987).
4. Kerckhove, C. *Géol. alpine* **45**, 5–204 (1969).
5. Cavelier, C. (coord.) *et al. Paléogène* in *Synthèse géologique du Sud-Est de la France.* Mémoire BRGM 125 (1984).
6. Kerckhove, C *C. r. hebd. Séance. Acad. Sci., Paris* **259**, 4742–4745 (1964).
7. Bodelle, J. thesis, Univ. Nice (1971).
8. Fry, N. *Geol. Soc. Lond., Spec. Publ.* **45**, 83–109 (1989).

The sublime and synaptic

John C. Marshall

Beauty and the Brain: Biological Aspects of Aesthetics. Edited by Ingo Rentschler, Barbara Herzberger and David Epstein. *Birkhäuser: 1988. Pp.325. SwFr.68.*

THEODOR Adorno's last book (*Aesthetic Theory,* 1970) opens with the line: "Today it goes without saying that nothing concerning art goes without saying". One should not, then, be unprepared for the existence of a volume with the subtitle *Biological Aspects of Aesthetics.*

No human culture is devoid of music and dance, the visual arts, poetry and story-telling. All normal children stomp unprompted to the band, scribble on the walls or doodle in the sand; devoted to our doggerel din, they demand the selfsame story 'til their parents' imperfect patience wears exceeding thin. No obvious utilitarian value attaches to such phenomena. The notion that the young are merely exercising their muscles and memories lacks conviction; the pleasure seems to be genuinely 'aesthetic' (although one might be hard pressed to provide rigorous corroboration) and intrinsic to our species.

In the rest of the animal world parallels are not easy to find. In Chapter 2 of *Beauty and the Brain,* Eibl-Eibesfeld mentions the 'tachist' paintings of Desmond Morris's chimpanzees and some interesting experiments (by Rensch) on monkeys, raccoons and birds. These latter studies in which the animals apparently showed preferences for symmetry and regularity are not sufficiently elaborated upon. It is true, of course, that birds sing, frogs croak and beavers build; bees prefer bright colours and primates beat their breasts. But these behaviours have a solid biological function; the attraction of mates or the repulsion of enemies; the provision of appropriate food and shelter.

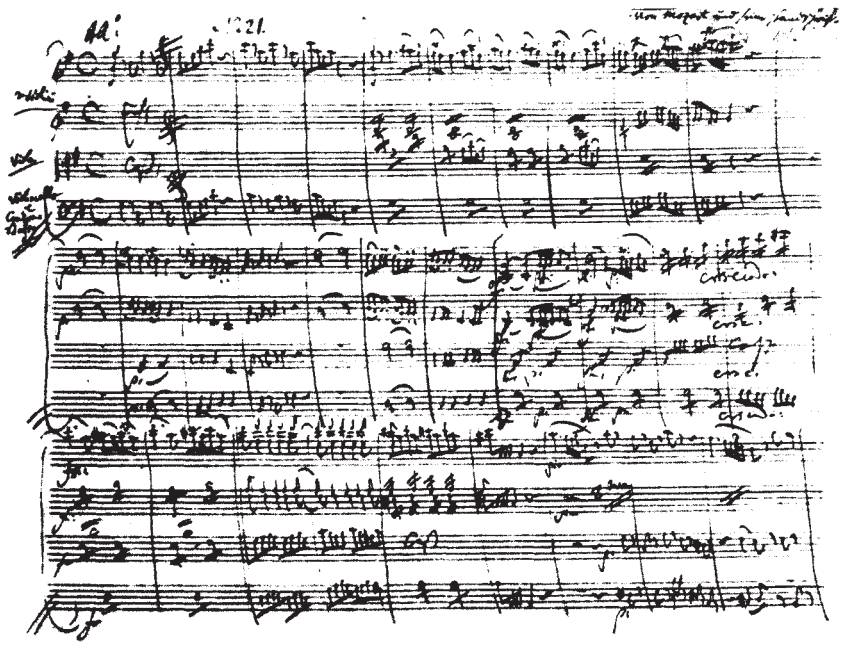
To credit other creatures with the capacity for aesthetic response is superfluous to Occam's razor. As far as we know, no Prince Charles of the spider-kingdom waves its legs in horror at the design of modern webs; no rodent that eats the green berries whilst avoiding the (poisonous) red ones would nonetheless collect the latter on the grounds of aesthetic superiority. Louder croakers may acquire fatter mates, but there is no firm evidence that painting abstract pictures for 12 hours a day will necessarily secure one a more attractive spouse (or pay the rent).

It is thus tempting to propose that the aesthetic sense is a distinct faculty of mind, represented physically in a dedicated brain-module; an autonomous, species-specific organ characterized by rule-governed creativity, the free play of which is engaged in for its own sake. This position is not dissimilar to current views of the organ of language, and is equally likely to produce apoplexy in those evolutionists who believe (as an article of faith) that all differences between related species are solely incremental. I exaggerate, of course, but not much. Indeed, Gregor Paul, one of the contributors to *Beauty and the Brain,* writes: ". . . if Chomsky's hypothesis that there exists a universal grammar is acceptable, then a similar hypothesis concerning universal aesthetic rules is even more convincing". Sadly, he does not develop this argument to the extent of telling us why Schönberg's fervent desire to be regarded as a high-class Tchaikovsky has not (yet) been fulfilled.

Beauty and the Brain derives from a series of symposia that took place in Bad Homburg between 1979 and 1983. The participants — "a group of neuroscientists, psychologists, anthropologists, philosophers, artists, musicians, and a poet" — met the magic number of seven times to lay the foundations for "the emerging field of *neuroaesthetics*". The book that has resulted from these deliberations is fascinating but very patchy.

As might be expected, the notion of hemispheric specialization looms large (although it is odd that the biological significance of artistic prodigies, either *idiots savants* or 'normal', is nowhere discussed). Regard and Landis show that preference judgements between simple forms differ as a function of the cerebral hemisphere to which the material is projected; Levy suggests that musicians and painters "have an unusual degree of bilateral representation of artistic skills, which permits superior interhemispheric integration". Grüsser, Selke and Zynda add a historical dimension by analysing the head position of 933 portraits from the fifteenth to the twentieth century; a distinct preference for showing the left side gradually decreases over the centuries. These chapters are not uninteresting although somewhat unbalanced. Data in search of theory or vice-versa is a general characteristic of the volume.

Thus Turner and Pöppel discuss metered poetry, to little artistic effect. They point out that two sounds must be (minimally) 0.03 seconds apart before their sequence can be perceived, that syllables are (typically) 0.3 seconds in duration, and that 3 seconds is "the length of the human present moment". But their account of meter is trivial and poorly integrated with their calculations: ". . . poetic meter serves a number of functions generally aimed at tuning up and enhancing the performance of the brain". Enhancing it for what? Epstein shows that across many



Notes for the aesthete — the first page of the autograph score of Mozart's *Eine Kleine Nachtmusik,* written in 1787. In some eyes, the score has a beauty of its own; some people, apparently, also enjoy the music when played.

cultures from Tibet to the Kalahari (via New Guinea and Venezuela) "tempos in musical performance change by low-order integral ratios". Both the biological and the aesthetic significance of the observation remain opaque. Siegfried analyses the space–time structure of the grasshopper dance performed by !ko-Bushmen. The pulse beat (indexed by the deepest flexion of the knees) is remarkably stable across extremely complex movement patterns but the author's interpretation is sloppy: "The rules provide a framework within which the dancer can cause shivers by trying new exciting movement combinations". You should see me do the tango!

The visual arts suffer much the same fate. Zollinger provides a good introduction to colour perception and naming, while Baumgartner writes lucidly on form perception. But neither explicitly relates the modular character of the visual system to the pictorial effects that can be manipulated in drawing and painting. The sole chapter that does make an effort to do so is by Rentschler, Caelli and Maffei. They demonstrate how reducing image information by band-pass and orientation-selective filtering can simulate the look of a Monet or a Feininger, or soften a Duccio into a Leonardo.

We now have a general model of the psychophysiology of the visual system: separate pathways for colour, different aspects of static form perception, and movement and depth, that are integrated into a unified percept. But why should (some ways of) de-unifying that percept by selectively enhancing, attenuating or distorting components of the *gestalt* produce interest, insight and pure pleasure? The editors of *Beauty and the Brain* did not invite "a recognized specialist in the field of art history or aesthetics" to their symposium. Future meetings may need to include contributions from those areas if neuroaesthetics is ever to replace the theory of the nine Muses. When actually listening to Mozart it is difficult not to feel that God decided to amuse herself by dictating a few piano concertos. And that choosing Amadeus as amanuensis was the best joke of all. □

John C. Marshall is in the Neuropsychology Unit, part of the Neuroscience Group at the Radcliffe Infirmary, Oxford OX2 6HE, UK.

On the road to Montreal

Philippa Lloyd

Ozone Crisis: The 15-Year Evolution of a Sudden Global Emergency. By Sharon L. Roan. *Wiley: 1989. Pp. 270. $18.95, £12.95.*

ONCE thought of as 'dream' compounds because of their chemical stability and lack of toxicity, chlorofluorocarbons (CFCs) first came under suspicion in 1974. In a paper published that year, Mario Molina and Sherwood Rowland of the University of California, Irvine, looked at what happened to them after they had been released into the atmosphere.

Because of their stability, CFCs reach the stratosphere unaltered but there they are broken up by short-wavelength ultraviolet radiation. Molina and Rowland deduced that this photodissociation would allow the release of free chlorine atoms which would then react with and ultimately destroy the ozone layer, the filter that protects the Earth from that same ultraviolet radiation. The worldwide production of CFCs at that time was almost a million tons a year, for use as refrigerants and in air-conditioning systems, as well as in spray cans, and the implications for the ozone layer were appalling.

Sharon Roan sets out to give a behind-the-scenes account of the controversy that ensued in the United States over the environmental safety, or otherwise, of CFCs. Writing very much from the viewpoint of Molina and Rowland, she vividly portrays the frustrations they faced in trying to alert the government and the chemical industry to their fears. But industrialists did not want theories, they wanted proof before they would be persuaded to take any kind of regulatory action.

The fight between industry and the atmospheric scientists was long and bitter, but in 1976 the Food and Drug Administration and the Environmental Protection Agency announced plans to phase out the use of CFCs in spray cans. The political climate was changing, however, and regulations to ban the use of non-aerosol CFCs failed to reach the statute book. In 1980, the new Reagan administration was reluctant to intervene in environmental issues — and in this case there was still a great deal of uncertainty over the extent to which CFCs were damaging the ozone layer. With science unable to resolve matters, lobbyists for industry were able to convince the legislators that any further regulations would be premature.

More evidence was needed to support Molina and Rowland's claims. In 1984 it arrived from an unexpected source. Joe Farman and his team at the British Antarctic Survey had for several years been monitoring ozone levels in the Antarctic stratosphere from their station at Halley Bay, and their data for 1977–1984 showed that a dramatic decline in ozone levels occurred in the southern spring. Farman's paper initially met with some scepticism, particularly as satellite data over the same area had not indicated any such loss of ozone. But it transpired that because the satellite instruments sometimes gave erroneous results, any readings below a certain value were flagged as 'bad' and thrown out. On reassessing the satellite data, the ozone depletion and its extent could easily be seen.

The reaction among the atmospheric science community was one of bafflement; although it had been predicted that CFCs might destroy ozone, nothing on the scale of the Antarctic 'ozone hole' had been forecast by any of the atmospheric models. The next step was to identify the cause, and in August 1986 13 scientists journeyed to the South Pole to take measurements in order to see what was going on. Uncertainties remained — there were both chemical and dynamical theories for the ozone hole — yet the following September delegates from around the world met in Montreal to sign a protocol proposing a 50 per cent cutback in the production of CFCs. A further expedition in August 1987 confirmed that CFCs were indeed the culprits and that November, after publication of the results, there were calls to strengthen the Montreal Protocol.

Since then, the Ozone Trends Panel has reported (in March 1988) that small ozone losses were occurring in the Northern Hemisphere, and in February of this year an international expedition to the Arctic announced that the chemical preconditioning necessary for ozone depletion had been found there too. NASA has just reported that this year the ozone hole over the Antarctic is as deep as that previously recorded (in 1987); it may turn out to be even deeper.

Just under half of Sharon Roan's text is devoted to events before 1984, the remainder to the discovery of the ozone hole and its ramifications — scientific, political and for industry. As a science writer on the *Orange County Register* in California, the author was well placed to cover the issues as they have developed over the past two decades. Her journalistic background betrays itself in the rather breathless prose, and the book is stronger on the politics and some of the personalities behind the ozone–CFC debate, rather than the science. But this is nonetheless a highly readable and informed account of its subject. □

Philippa Lloyd is an Assistant Editor of Nature.

Erratum

In the review of the Elsevier journal *Geomorphology* (*Nature* **341**, 362; 1989) price and frequency of publication should have been given as Dfl.298 and 4/yr, respectively. In the same issue (p.352) the price outside North America of *Chinese Journal of Biochemistry and Biophysics*, published by Allerton, should have read $280.

The correlation business

D. L. Dineley

Stratigraphy: Principles and Methods. By Robert M. Schoch. *Van Nostrand Reinhold: 1989. Pp.375. $41.95, £29.50.*

BORING stuff, stratigraphy — that seems to be a commonly held opinion in many quarters. To judge from the programme of the Twenty-eighth International Geological Congress in Washington last July, however, it is alive, well, and indeed so robust as to command an appreciable amount of the time available. Not only has a great deal of recent stratigraphic description revealed new and surprising phenomena, but the integration of stratigraphic studies with those on tectonics, neotectonics, sedimentation and conceptual modelling is providing new interpretations of geological processes and history. The stratigraphers of course have always maintained that it would and that writing historical geology — at least the last couple of thousand million years of it — is primarily dependent upon it.

The (simple) description of the rock strata of the crust of the Earth is now a highly technical business. All concerned have been keen to see established a common means of communicating and understanding knowledge of the stratigraphic record in different parts of the world. The North American Stratigraphic Code, produced by a North American Commission on Stratigraphic Nomenclature no less, has been closely followed in other parts of the world. An International Stratigraphic Guide followed (1977). An International Commission on Stratigraphy (ICS) had been set up and was in time recognized by the International Union of Geological Sciences (IUGS). In turn it has spawned a litter of subcommissions on the stratigraphy of the different systems, working groups on boundary problems and so on. The aim is to facilitate international uniformity of practice and ease of understanding, and the ICS guide is now being revised, as has been the American (1983).

Such activities as these bodies espouse may cut across national practices, and may appear legalistic and quibbling, but in the end their influence is for the good. They have to keep in mind technological progress as well as philosophical advances — magnetostratigraphy, ecostratigraphy, chronostratigraphy — and what *they* contribute to one of the prime aims of the whole wordy business. That aim is the global correlation of rock bodies and of the events and processes they have undergone. Telling the geological time is a complementary process and product.

Time lapse — medial and lateral moraines of the Barnard Glacier in Alaska, formed by deposition of debris from ice-flow. This photograph was reproduced from *The Nature of the Environment, Second Edition,* by Andrew Goudie, published by Basil Blackwell.

So much for the background, which goes to show why it is strange yet understandable that few authors have recently sought to produce a single volume that gives a global view of the principles of the discipline. Robert M. Schoch of Boston University has attempted just that task in the book under review. He has succeeded very well in what he describes as an introduction to a science in which there are often more questions than answers. To his 279 pages of text he appends a bibliography of 27 pages, plus the rather dated North American Stratigraphic Code (1983), a note on the definition and concept of geological–climate units, the timescale compiled by A.R. Palmer (1983) and the generalized lunar stratigraphy and geological timescale for the Moon.

The text is admirably concise and unambiguous. Schoch provides a general introduction to the science, describes the material basis of stratigraphy, and presents some of its conceptual foundations; the codes and conventions of stratigraphic nomenclature are dealt with, as are the different fields of physical stratigraphy, biostratigraphy and magnetostratigraphy, and the vexed topics of chronostratigraphy and 'geochronology'. (This last term is ambiguous, and the ICS is now avoiding it. The International Stratigraphic Guide uses it for the class of Period, Epoch, Age, chronozone while the Subcommission on Geochronology is concerned with radiometry leading to a geochronometic scale calibrating global chronostratigraphy.)

One may find minor points to disagree with in the book, but overall it is quite an achievement. The author is gratifyingly conversant with the history and development of his topic and this historical perspective is well treated. Keeping abreast with the developments of the past decade presents problems to every writer. Schoch, perhaps wisely, refrains from commenting on many of the most recent recommendations of the ICS and its subcommissions; for these, the reader will need to refer to the IUGS news magazine *Episodes*.

Something less than first class is the provision and choice of illustrations. They are insufficient in number, some are not very informative and others are simply not very good. Surprisingly, there is not a single graphic in the chapter on biostratigraphy and magnetostratigraphy.

For all that it is well written, this is a volume to be read in small spans. Most teachers and many researchers will find it very useful and informative, and one hopes (but rather doubts) that Earth scientists who are not primarily concerned with stratigraphic matters will browse in it. Part of the reason for that doubt is that the book is nicely produced but somehow 'low key': it is not eye-catching and in appearance is similar to a legal treatise. Stratigraphy is not boring, nor is this book — it just looks as if it is. □

D.L. Dineley is in the Department of Geology, University of Bristol, Bristol BS8 1RJ, UK.

Two diseases

Ann Harris and Martin Bobrow

Cystic Fibrosis. Edited by Peter Goodfellow. *Oxford University Press: 1989. Pp. 96. Pbk £15, $26.95*
The Fragile X Syndrome. Edited by Kay E. Davies. *Oxford University Press: 1989. Pp. 138. Pbk £15, $24.95*

THESE two slim, soft-covered books are the first in a new series, "Molecular Medicine", published by Oxford University Press. Each consists of five chapters by different authors, most of whom are actively engaged in research on the disease in question. In neither case do the editors appear to have imposed a uniform style, so that the individual 'mini-reviews' vary somewhat in style and quality.

Cystic fibrosis (CF) and fragile-X syndrome must be near the top of most people's list of fascinating monogenic human disorders. Cystic fibrosis is the most common serious recessive in populations of Northern European origin, with a gene frequency of about 2 per cent. The reason for this extraordinary frequency of a mutation which was, until recently, an effective genetic lethal, is unclear. Apart from this basic biological question, the disease is a considerable health problem in many countries. Electrophysiologists have gradually accumulated evidence that the basic defect lies in transmembrane ion transport. Within the past few weeks, CF has become the second major human disease locus to yield to pure 'reverse genetics', with the publication in *Science* (8 September, 1989) of the complete cDNA sequence and identification of the dominant disease-causing mutation. This achievement makes the book out of date, but renders it a useful standpoint from which to review the recent developments and those that will inevitably follow.

The first chapter is a straightforward account of the clinical features of CF, and the second deals with disease management. Against this clinical background, Chapter 3 is an excellent account of the complexities of ion transport systems and their abnormalities in CF. A basic introduction is followed by specific discussion of the sweat gland and airways' epithelia, which are known to exhibit abnormalities in CF. The fourth and fifth chapters (on molecular genetics and prenatal diagnosis) have been overtaken by events.

With a trait frequency of about 1 in 2,000 males, the fragile-X syndrome is also a formidable health-care problem and a tantalizing biological puzzle. It causes some degree of mental impairment in the great majority of male hemizygotes and a substantial minority of female heterozygotes. The clinical disorder behaves as a partially penetrant X-linked trait; the name reflects the presence of a site of preferential chromatid breakage at (or close to) the gene location of Xq27, manifest in a variable proportion of metaphase spreads from cultures treated with specific inducers or special media. Such site-specific cytogenetic manifestation of a monogenic disorder represents a unique interface between gene action, chromosome structure and clinical abnormality.

The opening chapter is again an account of the clinical aspects. It is followed by an excellent review of epidemiological studies of the disorder and its prevalence, and then a catalogue of some behavioural manifestations of fragile-X. The molecular geneticists' search for closely linked RFLPs in a condition of possible linkage heterogeneity, and with complexities in defining the phenotype, is described in Chapter 4, leaving an impression that the cloning of this particular gene may still be a little way off. A very competent review of the cytogenetics is for some reason left to the end of the book; this is a disorder that was first really defined at the cytogenetic level.

The format and pricing of these books make them accessible to a wide audience of students, and to both scientists and clinicians, each of whom is likely to learn something of the other's point of view. □

Ann Harris and Martin Bobrow are in the Division of Medical and Molecular Genetics, United Medical and Dental Schools of Guy's and St Thomas's Hospitals, Guy's Campus, London SE1 9RT, UK.

Our star

J.C. Brown

The Restless Sun. By Donat G. Wentzel. *Smithsonian Institution Press: 1989. Pp. 279. $27.95, £21.75*
The Sun: An Introduction. By Michael Stix. *Springer-Verlag: 1989. Pp. 390. DM.148, £48.50.*

As the Sun rises towards its peak magnetic activity, solar oscillation observers monotonically increase their efforts as too, it seems, do authors of solar physics books. The past few years have seen several new monographs on aspects of the subject at advanced student and introductory research level, notably those by Zirin (*Astrophysics of the Sun*), by Durrant (*The Atmosphere of the Sun*), and by Tandberg-Hanssen and Emslie (*Physics of Solar Flares*). Several others are on the way. All of these books have been valuable in meeting a long-standing need for updating recorded knowledge of various aspects of the field. But none has yet given coherent coverage of the whole subject rather than mainly the special interests of the authors.

The two volumes reviewed here are very different from one another in both level and style. Despite the generality of their titles, however, both once again fall short of providing thorough accounts of solar physics.

Wentzel's book is aimed at a general audience but is also announced as one that will be welcomed by specialists. The material is certainly easily readable and gives a clear impression of some areas of recent research. On the other hand, I found the style and presentation somewhat banal and the coverage in many areas thin and rather arbitrary. The general tone is reflected in the reference on the dust-jacket to the coronal temperature in degrees Fahrenheit. Such units are not likely to be welcomed by the general scientific reader, let alone the specialist, any more than are the trite comments to the effect that work in this field "needs only modest equipment" and "developing countries have the advantage of lower labor costs". Although such a general book cannot be expected to be comprehensive, it is disappointing to find that most areas of the subject are presented, and referenced, only in terms of very specific observations and theories, some outmoded and some speculative. On the other hand, the monochrome illustrations are too poor and the colour plates too few greatly to enhance a coffee table.

In contrast, Stix's book is aimed at advanced students. It is a careful, well-written, scholarly text, giving thorough coverage of a considerable number of selected areas, both observational and theoretical, and from physical and mathematical viewpoints. As the author himself emphasizes, the areas covered well centre on his own interests in the solar interior (oscillations, convection and magnetism), and there are sound chapters on observational tools and problems, and on the lower atmosphere. Thereafter there is a brief account of the solar chromosphere, corona, wind and flares. Unfortunately this final chapter is very restricted in both subject matter and references, giving a distorted picture of the main directions of much recent research in these topics. Nevertheless, within its limits *The Sun* is a solid piece of work.

At the advanced level, solar physics is most comprehensively covered by *The Sun as a Star*, edited by S. Jordan, and the three-volume set *Physics of the Sun*, edited by P.A. Sturrock. But these collections of articles are dating fast, and there remains a gap in the literature for a scholarly monograph covering all of the field. □

J.C. Brown is in the Department of Physics and Astronomy, University of Glasgow, Glasgow G12 8QW, UK

Brief history of the Phobos mission

R. Z. Sagdeev & A. V. Zakharov

The spacecraft Phobos 1 and 2 were launched on 7 and 12 July 1988. The mission was to have three stages: investigations of the Sun and interplanetary space during the flight from Earth to Mars; studies of both Mars and Phobos during the orbit of the spacecraft around Mars; and studies of Phobos as the spacecraft approached to within 50 m of its surface. Contact with Phobos 1 was lost on 1 September 1988 and contact was also lost with Phobos 2 just before the third stage of the mission. This article introduces fourteen papers which cover the results of the second stage of the mission.

NO spacecraft were sent to Mars for 13 years after the Viking 1 mission. This 'moratorium' was broken in 1988 by the launch of two Soviet spacecraft, Phobos 1 and Phobos 2. The main destination of the mission was not Mars, but Phobos, one of its two moons. This reflected the growing interest in the minor bodies of the Solar System. After the success of the Vega, Giotto and Suisei missions to comet Halley, it was natural to 'try' an asteroid or a little moon. From this point of view, Phobos was an ideal choice, as in its spectral characteristics, it resembles a real asteroid, but is much easier to approach, using the martian gravitational field to manoeuvre.

Initial information about the martian moons was based on results from the Mariner 9 and the Viking missions. The minimum distance at which Viking 1 approached Phobos was ~100 km. The Phobos spacecraft was to have come as close as ~50 m to Phobos which would have made it possible to obtain high-spectral-resolution television images of the surface of Phobos, to study the elemental and mineralogical composition of the regolith, the internal structure of Phobos, and to put landers on the moon's surface.

It is safe to say, however, that Mars itself was not exhausted as an object of exploration. Moreover, Mars will hopefully play a central part in the long-term Soviet programme of Solar System exploration.

Besides studies of Phobos, and the martian surface, atmosphere and plasma environment of Mars, both of the Phobos spacecraft were to make observations of the Sun and interplanetary space during the flight from Earth to Mars[1].

A large international team took part in the preparation of the Phobos project. The onboard scientific payload was prepared by scientific teams from Austria, Bulgaria, Hungary, East and West Germany, Ireland, Poland, Soviet Union, Finland, France, Switzerland, Sweden, Czechoslovakia and the European Space Agency (ESA). The US Deep Space Network participated in the ballistic support of mission operation.

Spacecraft and scientific payload

Phobos (Figs 1 and 2) is a next-generation spacecraft after those of the Venera-type, last used in the Vega mission (1984–86). The Phobos spacecraft, developed in the Babakin Centre, was planned as a base spacecraft for the next generation of interplanetary missions[2]. It took eight years to complete the technical and scientific preparation of Phobos, and the schedule, which was altered drastically several times, was to be as follows: (1) The spacecraft would reach the vicinity of Mars and be put in an elliptical orbit; (2) several successive corrections would bring the orbit closer to that of Phobos. During this period Mars and Phobos would be studied by remote methods, the latter up to ~100 km; (3) in the culmination, the spacecraft would approach the surface of Phobos, to an altitude of 30–50 m for a short time, release two small landers and perform two 'active' probing experiments. One of the latter involved the use of laser beams to evaporate small samples of material and analysis of the samples by sensitive mass spectrometers. Unfortunately this, the most novel part of the programme was not realized.

The Phobos mission involved the launch of two spacecraft, Phobos 1 and Phobos 2. The total pre-launch weight of each spacecraft was 6,200 kg, of which 3,600 kg was the weight of the propulsion system, which was to be separated from the spacecraft after the most power-consuming man- oeuvres of the spacecraft were performed in setting up the reference orbit.

The contents of the scientific payload aboard Phobos 1 differed slightly from that of Phobos 2. The total weight of scientific payload aboard each was ~500 kg (with structural elements and cables). In accordance with the scientific aims of the mission, the instrumentation may be classified as planetary, plasma and solar. Table 1 presents the scientific instruments of the orbiter which provided the 'planetary' data of the mission, that is, data from the exploration of Phobos, the martian surface and atmosphere. All these instruments were mounted within the spacecraft so that their sensitive elements were oriented in

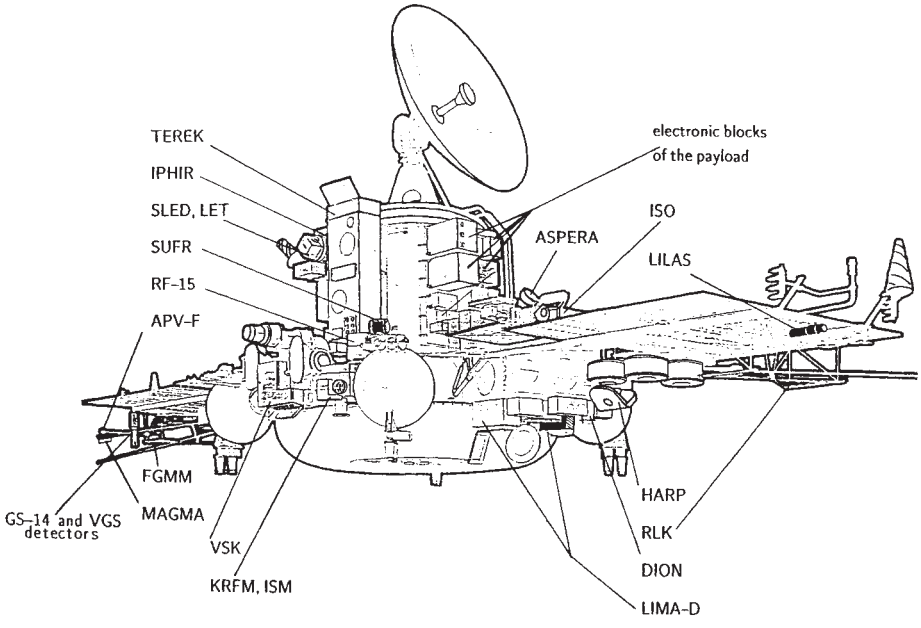


FIG. 1 Arrangement of scientific instruments aboard Phobos 2.

an anti-sunward direction (with the exception of ISO). The spacecraft was to change its orientation to point the instruments towards Phobos or Mars while on the martian orbit. One of the Phobos landers was capable of remaining at the surface for a long time and contained instruments for celestial mechanics experiments, analysis of regolith and for television imaging of landing places. The other lander, called a 'hopper', could jump over the surface to study the composition of the Phobos soil in several places.

The 'plasma' part of the programme concentrated on the exploration of the magnetic field and plasma environment of Mars, measurements of the parameters of the interplanetary medium, the Sun and cosmic rays. A list and the main characteristics of the instruments are given in Table 2. Table 3 presents the solar-package instrumentation and γ-burst experiments.

The project and the scientific programme

The Phobos 1 and Phobos 2 spacecraft were launched on 7 and 12 July 1988, respectively, by Proton rockets from the Baikonur cosmodrome. On 16 and 21 July the first correction of the trajectory was made. On 1 September, Phobos 1 did not establish radio contact during the planned session with Earth and did not answer multiple radio requests. The analysis of up-link commands sent during the previous communication session, that is on 28 August, showed that an erroneous command had been given which switched off the thrusters of the attitude control system. Phobos 1 changed its orientation and solar panels were turned away from the Sun. When this mistake was discovered, on 1 September, the power supply available on board was not enough to react to powerful radio signals from the Earth. Thereafter the mission had only Phobos 2 to carry out the scientific programme.

On 29 January 1989, a retarding impulse was applied ($\Delta v = 815.1\ \mathrm{m\,s^{-1}}$) and Phobos 2 was transferred from the Earth-to-

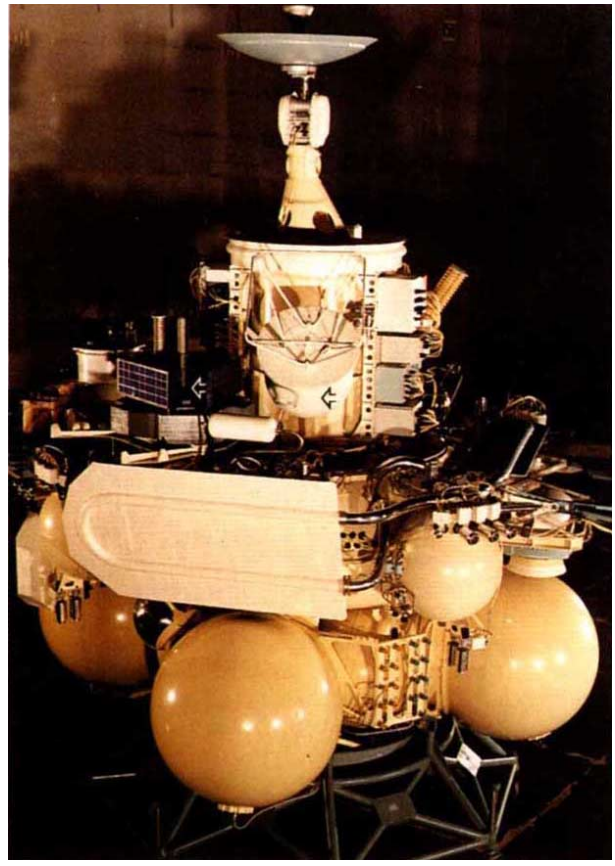


FIG. 2 Phobos 2: white arrow, long-term autonomous lander; black arrow, 'Hopper'.

TABLE 1 Scientific payload for planetary studies

Instruments	Spacecraft	Mass (kg)	Main features	Cooperation	References
Videospectrometric system (VSK)	1,2	51.7	TV camera, $\lambda = 0.4$–$0.6\ \mu$m, $\Delta\theta = 4$ arcmin; $\lambda = 0.8$–$1.1\ \mu$m, $\Delta\theta = 4$ arcmin; $\lambda = 0.4$–$1.1\ \mu$m, $\Delta\theta = 0.7$ min of arc; image size, 288×505 pixels. Spectrometer, 36 spectral bands within 0.5–1.0 μm.	Bulgaria, GDR, USSR	3
Infrared radiometer/spectrometer (KRFM)	2	19*	Spectrometer, $\lambda = 0.3$–$0.9\ \mu$m, $\Delta\lambda = 50$ nm, $\Delta\theta = 16$ arcmin; radiometer, $\lambda = 5$–$67\ \mu$m, 6 bands, $\Delta\theta = 30$ arcmin	France, USSR	4
Infrared spectrometer (ISM)	1,2		Spectral range, $\lambda = 0.8$–$3.5\ \mu$m; spectral resolution; $\Delta\lambda = 20$ nm; viewing angle: $\Delta\theta = 12$ arcmin	France, USSR	5
Scanning infrared radiometer (Thermoscan)	2	28	Spectral range, $\lambda = 8.5$–$12\ \mu$m and $\lambda = 0.6$–$0.95\ \mu$m, viewing angle, $\Delta\theta = 0.9$ arcmin; scanning angle: 6.1°; scanning rate, 1 line $\mathrm{s^{-1}}$	USSR	6
γ-Emission spectrometer (GS-14)	1, 2	11.5	Energy range, 0.1–10 MeV; spectral resolution, 12%; detector CsI(T1), $V = 785\ \mathrm{cm^3}$	USSR	7
Neutron detector (IPNM)	1	12	Energy range, 0.025–5 MeV; time resolution, 100 ms	USSR	8
Radar system (RLK)	1, 2	40.9	Frequency, 5, 130, 500 MHz	USSR	9
Laser mass-spectrometric analyser (LIMA-D)	1, 2	81.5	Mass range, 1–200 AMU, $M/\Delta M = 150$; energy density $10^9\ \mathrm{W\,cm^{-1}}$	Austria, Bulgaria, Czechoslovakia Finland, FRG, GDR, USSR	10
Secondary-ion mass-analyser (DION)	1, 2	23.9	Mass range, 1–60 AMU, $M/\Delta M = 100$; $M/\Delta M = 100$; ion beam current, 2 mA	Austria, Finland, France, USSR	11
Optical radiation spectrometer (ISO)	1, 2	18.1	Spectral bands, $\lambda = 255$ nm, $\lambda/\Delta\lambda = 300$ (O_3); $\lambda = 760$ nm, $\lambda/\Delta\lambda = 10^5$ (O_2); $\lambda = 950$ nm, $\lambda/\Delta\lambda = 10^5$ (H_2O); $\lambda = 3.7\ \mu$m, $\lambda/\Delta\lambda = 10^3$ (HDO)	France, USSR	12, 13

* Total mass of the KRFM and ISM instruments.

TABLE 2 Scientific payload for plasma studies

Instruments	Spacecraft	Mass (kg)	Main features	Cooperation	References
Magnetometers			Measurement range 100 nT	Austria, USSR, GDR, USSR	14
MAGMA	1, 2	4.5			
FRMM	1, 2	5.1			
Plasma wave analyser (APV-F)	1, 2	7.7	Frequency range, $E_z = 10^{-3}$–(2×10^5) Hz; $B_y = 10$–(4×10^4) Hz; $n_iV = 10^{-3}$–(5×10^2) Hz;	ESA, Czechoslovakia, USSR	15
Scanning energy-mass spectrometer (ASPERA)	1, 2	8, 7	Energy range, ions, 0.5eV q–25 keV q^{-1}; electrons, 0.5 eV–50 keV; geometric factor, $GxE = 2\times10^{-4}$ cm^2 sr keV (ions)	Finland, Sweden, USSR	16
Energy-mass charge spectrometer (SOVIKOMS	1, 2	16*	Energy range, 50^{-1}–30 keV q^{-1}	Austria, FRG, Hungary, USSR	1
Proton and α-particle spectrometer (TAUS)	1, 2		Energy range, 0.15–6 keV q^{-1}; geometric factor, $G = 10^{-3}$ cm^2 sr	Hungary, USSR	17
Ion and electron spectrometer (HARP)	1, 2	16.2†	Energy range, 0.5–1,000 eV	Hungary, USSR	18
Energetic charged-particle spectrometer (SLED)	1, 2		Energy range, 0.02–3.4 MeV; geometric factor, 1 cm^2 sr	FRG, Hungary, Ireland, USSR	19

* Total mass of the VOSIKOMS and TAUS instruments.
† Total mass of the HARP, SLED and LET instruments.

Mars trajectory to the elliptical equatorial orbit of a martian artificial satellite. Table 4 summarizes the parameters of the elliptical orbits. On 12 February, after four turns around Mars, the orbit was corrected ($\Delta v = 114.6$ m s^{-1}) so that its pericentre came closer to the orbit of Phobos. On 18 February one more retarding impulse was applied ($\Delta v = 722$ m s^{-1}) and the spacecraft's orbit became circular, almost the same as that of Phobos although somewhat higher. Table 4 lists the parameters of this circular orbit. This manoeuvre completed the circular-orbit formation, and the self-contained propulsion system was jettisoned. The spacecraft's motion was then controlled by low-thrust engines.

After two intermediate corrections of the circular orbit were made on 7 and 15 March ($\Delta v = 37.6$ m s^{-1} and 1.0 m s^{-1}, respectively) the spacecraft was transferred to a quasi-synchronous orbit ($\Delta v = 39.7$ m s^{-1}) on 21 March (orbiting around Mars practically in synchrony with Phobos). For the coordinate system fixed with the three-axis-stabilized spacecraft, this led to Phobos spinning at the anti-solar side of the spacecraft at a distance of 200–600 km. The next stage involved transferring Phobos 2 to an orbit closer to Phobos and then its controlled departure from this orbit to approach the satellite. Figure 3 presents schematically the Phobos 2 orbits around Mars. The Phobos surface image (see Fig. 4) taken by the Phobos 2 television system has marked sites for landers, as well as the region over which the spacecraft should drift at the closest approach.

On 27 March 1989, after television imaging of Phobos during which Phobos 2 changed its orientation so that its field of view (FOV) pointed at Phobos, we lost radio contact with the spacecraft. Four hours later, after unsuccessful attempts to set up radio contact, a weak signal was received with the last message from Phobos 2: the spacecraft was spinning in the off-design mode, extremely adverse in terms of power supply for the spacecraft systems, and was unable to support reliable radio communication with Earth. Further attempts to communicate with Phobos 2 were ineffective. Although this time, controller error was not the cause, the spacecraft loss was again due to failure in orientation control, which quickly led to energy exhaustion.

We are still awaiting the final verdict as to how Phobos 2 was lost, but several suggestions have been made. One is that orientation could be lost because of on-board computer malfunction. The contractor (Babakin Center) raised the question of possible impact between the spacecraft and large dust particles, the concentration of which should be, in principle, at a maximum around the orbit of Phobos because of regolith secondary particles knocked out by micrometeorite and meteorite impacts.

Because this problem might be of particular interest for future missions we consider a simple model of a dust torus on orbits near Phobos. For estimation purposes it is reasonable to proceed by analogy with the regolith particle knocking-out rate on the moon. However, the majority of secondary particles knocked out from the lunar surface return with free-fall time because of gravity. For Phobos, which has insignificant gravity (with the escape velocity slightly over $v = 10$ m s^{-1}), we consider, for simplicity, that all the material knocked out from the surface goes into orbit around Mars. The resurface rate is assumed, by the order of magnitude, to be close to that on the Moon (10^{-7} cm yr^{-1}). Sooner or later these particles, depending on the characteristics of their orbits, should come into contact with Phobos. So, at the equilibrium of Phobos with the dust torus, the amount of secondary particles deposited on the surface

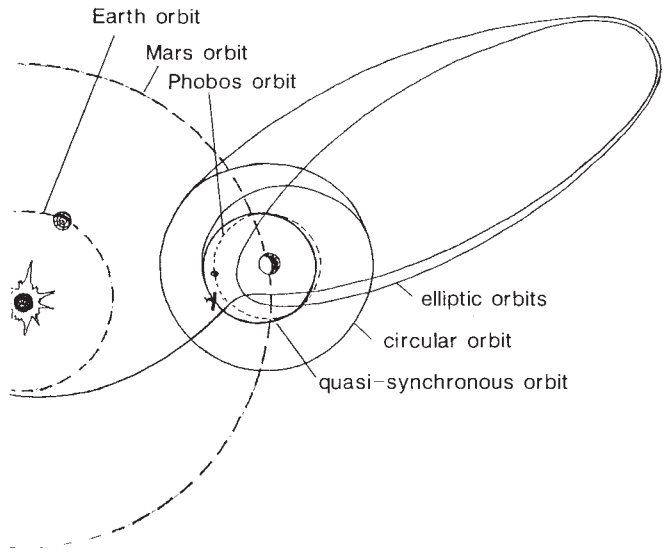


FIG. 3 Scheme of the Earth, Mars, Phobos orbits and the orbits of Phobos 2 (not to scale).

TABLE 3 Scientific payload for studies of the Sun

Instrument	Spacecraft	Mass (kg)	Main features	Cooperation	References
Solar photometer (IPHIR)	1, 2	6.3	Wavelengths, *V* 335,500,862 nm; band width, 5 nm; resolution, $\Delta l/l < 10$	ESA, France, Switzerland USSR	20
Solar telescope/coronograph (TEREK)	1	36	Spectral range, 0.5–6.0 nm, 17–18 nm, 30.4 nm, 400–600 nm	Czechoslovakia, USSR	21
X-ray photometer (RF-15)	1, 2	6	Energy range, 2–4 and 4–8 keV	Czechoslovakia, USSR	1
Ultrasound-spectrometer (SUFR)	1, 2	3.6	Wave range, 0.3–2.5; 0.3–12; 121.6; 130 nm	USSR	1
γ-burst spectrometer (LILAS)	1, 2	5.9	Energy range, 3 keV–1 MeV	France, USSR	22, 23
γ-burst spectrometer (VGS)	1, 2	2.3	Energy range, 0.1–10 MeV	France, USSR	1

should, on average, be the same as those that left it. Without going into details of stochastic diffusion of these particle orbits (according to calculations made by Dr G. M. Zaslavsky's team), which ensures mixing for 1 yr, it is easy to evaluate the probability of a spacecraft's impact with dust particles occurring within the dust torus. To find the upper limit, we suppose that the secondary mass of knocked-out regolith is converted to 10 g mass 'bullets' (this mass was conditionally considered as a danger for the spacecraft). With the regolith density 2 g cm^{-3}, the surface of Phobos loses (and hence, takes back) $\sim 10^5$ g of particles per year. The spacecraft, with an effective area of $\sim 30\ m^2$, will, on average, collide once every 300 yr. It should again be mentioned that the probability of the spacecraft colliding with a particle of mass ~ 10 g is, in fact, many orders lower.

The similar conclusion that the probability of spacecraft impact with a dangerous particle is extremely low, was made by the Viking team (B. Murray, personal communication).

Conclusion

The failure of the Phobos spacecraft before completion of the mission invites a discussion on the interaction of scientists with industry (customers with constructors) to ensure reliability and survival of a mission. Glavkosmos in the Soviet Union (the prime contractor) has promised to change its approach to future missions (would this be real '*perestroika*'?). The scientific and engineering team must discuss technical problems in the early stages of project preparation.

The results of measurements made during the flight from Earth to Mars (52 days for Phobos 1 and 200 days for Phobos 2) as well as in orbits around Mars (57 days for Phobos 2) have their own significance; results from the second stage of the mission will be discussed in the papers which follow. □

Note added in proof: Since the paper was submitted, it has been verified that Phobos 2 was lost because of an on-board computer malfunction.

R. Z. Sagdeev, Scientific Director of the Phobos Project, and A. V. Zakharov, Phobos Project Scientist, are at the Space Research Institute, Profsoyuznaya 84/32, 117810 Moscow GSP-7, USSR.

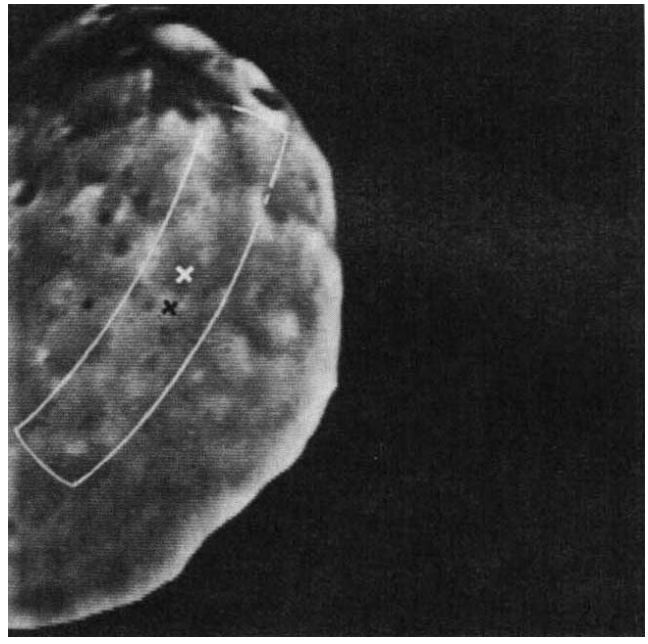


FIG. 4 The Phobos-surface image taken by the Phobos 2 TV system. The sites at which the spacecraft should be pointed while approaching the surface are marked with crosses. The region over which the spacecraft should drift with due account of probable deflection of its trajectory is also shown.

1. Sagdeev, R. Z., Balebanov, V. M. & Zakharov, A. V. *Soviet Sci. Rev. E. Astrophys. Space Phys.* **6**, 1–60 (1988).
2. Kremnev, R. S., Rogovskyi, G. N. & Sukhanov, K. G. *Sci. USSR* **1**(43), 2–8 (1988).
3. Avanesov, G. A. *et al. Nature* **341**, 585–587 (1989.
4. Ksanfomalit, L. V. *et al. Nature* **341**, 588–591 (1989).
5. Bibring, J.-P. *et al. Nature* **341**, 591–593 (1989).
6. Selivanov, A. S. *et al. Nature* **341**, 593–595 (1989).
7. Surkov, Yu. A. *et al. Nature* **341**, 595–598 (1989).
8. Surkov, Yu. A. *et al. Phobos, Space Res. Inst.* 277–286 (1988).
9. Armand, N. A. *et al. Phobos, Space Res. Inst.* 313–333 (1988).

TABLE 4 Orbital parameters

Orbit no.	1	2	3	4	5	Circular orbit
			Elliptical orbits			
Flight day	204	205	206	207	210	
Data of pericentre passage day/month (1989)	01/02	05/02	08/02	11/02	14/02	
Time (UT) pericentre passage	18:39.37	00:15.47	05:51.58	11:28.12	23:38.50	
T (h)	79.11	79.12	79.13	79.14	86.66	8.013
α (km)	44,479.1	44,482.4	44,485.7	44,489.1	47,266.1	9,668
ε	0.905210	0.904406	0.904604	0.904801	0.792630	0.0130
ω (deg)	69.81	70.42	70.99	71.52	131.01	328
I (deg)	0.866	0.865	0.864	0.863	0.876	1.26
Ω (deg)	198.596	198.165	197.779	197.441	143.332	297.27
H_π (km)	866.9	858.2	850.0	841.3	6,407.5	6,144.8

10. Sagdeev, R. Z. *et al. Phobos, Space Res. Inst.* 220-230 (1988).
11. Hamelin, M. *et al. Phobos, Space Res. Inst.* 245-260 (1988).
12. Blamont, J. E. *et al. Nature* **341,** 600-603 (1989).
13. Krasnopolsky, V. V. *et al. Nature* **341,** 603-604 (1989).
14. Riedler, W. *et al. Nature* **341,** 604-607 (1989).
15. Grard, R. *et al. Nature* **341,** 607-609 (1989).
16. Lundin, R. *et al. Nature* **341,** 609-612 (1989).
17. Rosenbauer, H. *et al. Nature* **341,** 612-614 (1989).
18. Shutte, N. M. *et al. Nature* **341,** 624-616 (1989).
19. Afonin, V. V. *et al. Nature* **341,** 616-618 (1989).
20. Fröhlich, C. *et al. Proc. Symposium Spain,* 26-30 Sept. 1988, ESA SP-286 (1988).
21. Sobelman, I. I. *et al.* Preprint 241 (Labedev Institute, Moscow 1988).
22. Mitrofanov, I. G. *et al.* Preprint 1450 (Space Research Institute, Moscow 1988).
23. Mitrofanov, I. G. *et al.* Preprint 1497 (Space Research Institute, Moscow 1988).

ACKNOWLEDGEMENTS. We thank all the scientists and engineers involved in the project.

Television observations of Phobos

G. A. Avanesov*, B. I. Bonev†, F. Kempe‡, A. T. Bazilevsky§, V. Boycheva†, K. N. Chikov||, M. Danz¶, D. Dimitrov†, T. Duxbury#, P. Gromatikov†, D. Halmann‡, J. Head, V. N. Heifets*, V. Kolev†, V. I. Kostenko*, V. A. Kottsov*, V. M. Krasavtsev||, V. A. Krasikov*, A. Krumov†, A. A. Kuzmin*, K. D. Losev||, K. Lumme††, D. N. Mishev†, D. Möhlmann¶, K. Muinonen††, V. M. Murav'ev*, S. Murchie**, B. Murray‡‡, W. Neumann‡, L. Paul‡, D. Petkov†, I. Petuchova‡, W. Pössel‡, B. Rebel‡, Yu. G. Shkuratov§§, S. Simeonov†, B. Smith||||, A. Totev†, Yu. Uzunov†, V. P. Fedotov*, G.-G. Weide‡, H. Zapfe¶, B. S. Zhukov* & Ya. L. Ziman***

* Space Research Institute, Profsoyuznaya 84/32, 117810 Moscow, GSP-7, USSR
† Space Research Institute, Moskovskaya 6, Sophia, Bulgaria
‡ Zentralinstitut fur Kybernetic und Informations Prozesse, Kurstraße 3, Berlin, DDR
§ Institute for Geochemistry and Analytical Chemistry, Kosigina 19, 117975 Moscow, GSP-1, USSR
|| Institute for Fine Mechanics and Optics, Leningrad, USSR
¶ Institut für Kosmosforshung, Rudower Chausse, 1199 Berlin, DDR
# Jet Propulsion Laboratory, Pasadena, California 91109, USA
** Brown University, 1846 Providence, Rhode Island 02912, USA
†† University of Helsinki, Tahtitorninmaki, SF-00130 Helsinki 13, Finland
‡‡ California Institute of Technology, Pasadena, California 91125, USA
§§ Harkov University, 310077 Harkov-77, USSR
|||| University of Arizona, Lunar and Planetary Laboratory, Tucson, Arizona 85721, USA

IN February and March 1989 the Phobos 2 spacecraft took 37 television images of the moon Phobos from a distance of 190–1,100 km. These images complement the data from the Mariner 9 and Viking missions by providing higher-resolution coverage of a large region west of the crater Stickney (70–160° W) and by providing disk-resolved measurements of surface brightness at a greater range of phase angles and wavelengths. The data are being used to update the three-dimensional model of Phobos, to provide improved determinations of its density and orbital dynamics and to study its surface colour, composition and texture. Here we present preliminary findings, which include different integrated photometric behaviour in visible and near-infrared bands, observation of a region immediately west of Stickney which is relatively free of large grooves, the prevalence of bright rims on grooves and younger craters and low bulk density.

The videospectrometric system (VSK) Fregat was developed cooperatively by Bulgaria, East Germany and the Soviet Union. VSK combines a three-channel television (TV) camera, a spectrometer, and a memory with a capacity >1.5 Gbit. The TV camera had two short-focal-length ($f = 18.5$ mm) channels and one longer-focal-length ($f = 100$ mm) channel with angular resolutions of 3.3×4.5 arcmin and 0.62×0.83 arcmin, respectively. The spectral sensitivities of the three TV channels are shown in Fig. 1. The TV camera and spectrometer had charge-coupled device (CCD) detectors with photosensitive sections and memory sections of 288×505 pixels, and used an 'electronic shutter' technique.

Because of the premature loss of the Phobos 2 spacecraft, the TV observations of Phobos were limited to 37 images of Phobos and Phobos against Mars, including 13 higher-resolution images in channel 2 which cover most of the surface of Phobos except for that obscured by Mars. Over two-thirds of Phobos had been viewed by Viking at a spatial resolution of 50 m pixel^{-1} or better. However, the area to the west of Stickney (70–160° W) had previously been viewed at a very oblique angle and at a spatial resolution of a few hundred metres per pixel. VSK imaged this area at low emergence angles and at a spatial resolution of 40–80 m pixel^{-1} (Fig. 2*b*, *d*, *e*). Because of the generally low phase angles the topography of this area is still not seen very clearly, but craters and grooves can be recognized. Comparison with Viking-based maps[1,2] shows that several new grooves can be identified in the area 130–180° W (Fig. 2*e*); imaging under the same conditions failed to reveal new grooves closer to Stickney (Fig. 2*d*), an area already known from Viking images to be relatively free of larger grooves. Either large grooves did not form here or they were buried by ejecta from Stickney.

The control network of Phobos derived from Viking data[3] was found to be generally consistent with VSK images. However, deviations of up to 0.5–1 km were observed when limb profiles of Phobos were compared with those derived from the Phobos topographic map[4].

In images taken at phase angles of 7–8° (Fig. 2*d*, *e*), many craters and grooves are observed to have on their rims and walls, material that is 20–30% brighter than average for Phobos surface materials. At moderate phase angles (Fig. 3*b*, *c*, *f*) the difference in brightness disappears, implying that this material has a higher normal albedo as well as a higher phase coefficient at small phase angles. This effect had also been observed on Viking images[1,5,6] and had been interpreted as resulting from variations in surface texture.

Alternatively, on dark bodies like Phobos both a higher normal albedo and a higher phase coefficient can be explained by the diffraction mechanism of opposition surge[7], as a higher particle single-scattering albedo possibly resulting from less regolith darkening and maturation (the common shadow mechanism implies decrease of the phase coefficient with increase of single-scattering albedo). The lesser maturity of regolith on the crater

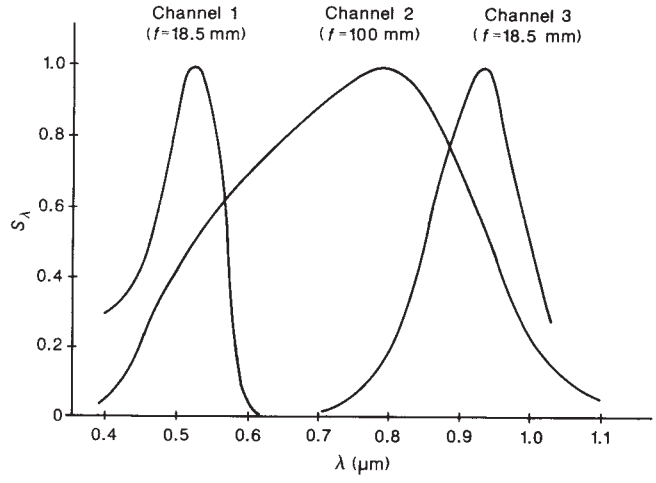
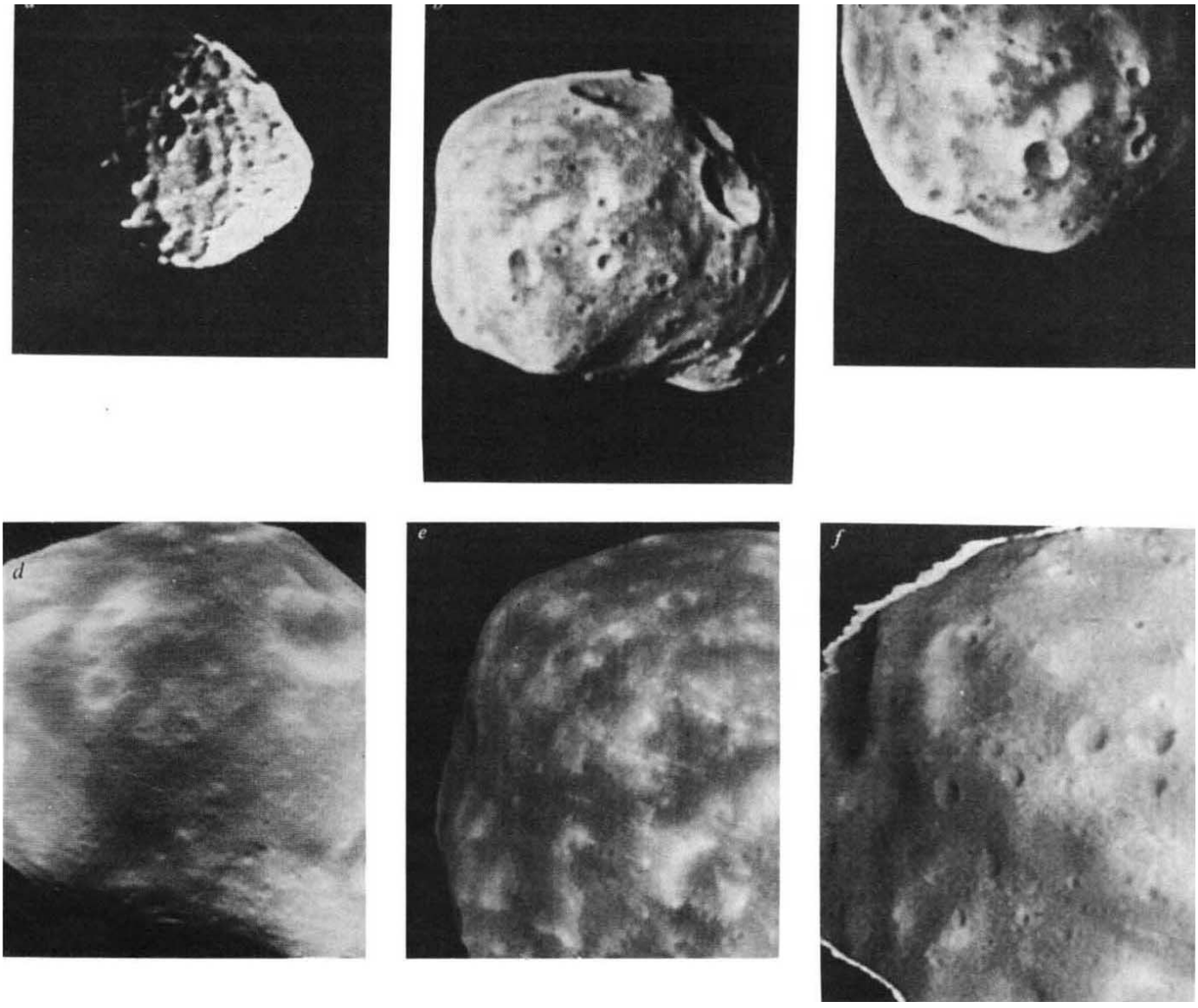
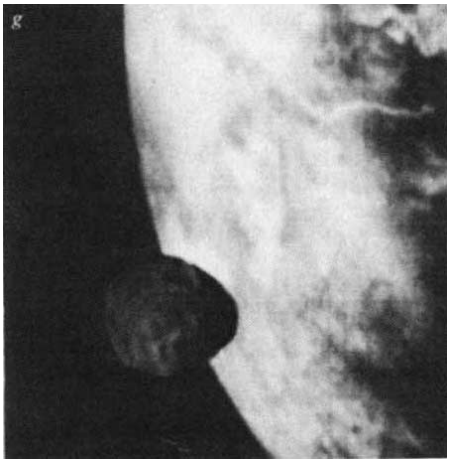


FIG. 1 Spectral sensitivities of the VSK TV channels. S_λ, relative spectral sensitivity; λ, wavelength (μm); *f*, focal length.





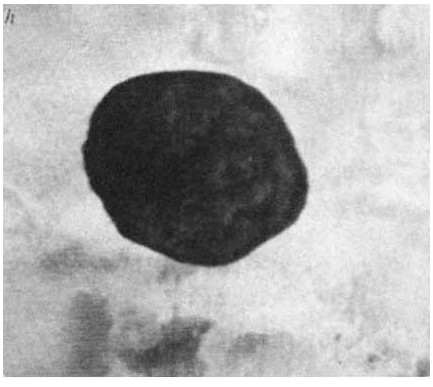


FIG. 2 Enhanced VSK images of Phobos and Phobos against Mars (processing by IKI, USSR). For each image, the channel, sub-spacecraft latitude, and longitude, spacecraft–Phobos distance, pixel size (on Phobos) and phase angle, respectively, are as follows: *a*, 2, 4°, 254° W, 958 km, 173 m, 74°; *b*, 2, 7°, 134° W, 439 km, 79 m, 24°; *c*, 2, 19°, 213° W, 359 km, 65 m, 19°; *d*, 2, −1°, 111° W, 282 km, 51 m, 8°; *e*, 2, 0°, 149° W, 203 km, 37 m, 7°; *f*, 2, 1°, 179° W, 191 km, 34 m, 18°; *g*, 3, 16°, 162° W, 335 km, 326 m, 25°; *h*, 1, 1°, 181° W, 192 km, 187 m, 19°. At large phase angles (*a*) the complicated relief of Phobos is seen very clearly with the surface saturated with craters. At low phase angles (*d*, *e*) the brightness only slightly depends on surface slopes, and optical characteristics and texture of the surface can be analysed more reliably.

and groove rims may be a result of their younger age or of sliding of surface regolith off steeper slopes, perhaps driven, under the low gravity of Phobos, by impact-generated seismic vibrations or by diurnal surface-temperature variations.

To test these hypotheses, the relationship of photometric properties of crater rims and walls with the degradation state of the crater was investigated. This was done by mapping and classifying craters of diameter >1 km within the area observed at low phase angles in the combined Viking and Phobos 2 data sets, extending from ~50–70° N to 50° S and from 40° to 300° W. Each crater was classified into one of four morphological groups previously defined in ref. 3, although assignments of individual craters to different groups were revised on the basis of the new VSK images for the area 70–160° W. In addition, each crater

TABLE 1 Extent of bright material in different crater classes

Crater group	Number of craters	Extent of bright material: More than half the rim (%)	Less than half the rim (%)	Absent (%)
Fresh	3	100	—	—
Partly degraded	28	57	15	28
Degraded	8	12	38	50
'Ghost' craters	6	—	33	67

was classified as one of three types based on the presence and the continuity of bright material along the rim. The results, summarized in Table 1, clearly show that bright rims and walls are prevalent on fresher craters and disappear as the craters are degraded, consistent with the occurrence of bright rims on younger craters with less mature regolith.

Higher brightness at low phase angles is also shown by rim materials of grooves, although generally not by groove floors. Several mechanisms by which grooves could be formed have been proposed, and testable predictions have been made about the physical properties of grooves[8]. The models are: (1) the 'fracture and drainage' model[2] which proposes exposure of fresher material by sliding of regolith from groove walls and floors; (2) the 'fracture and degassing' model[1,2], which proposes an accumulation on groove rims of material from depth that may have a composition or texture different from that of surface material; and (3) the 'ejecta re-impact' models[9,10], whereby groove-floor material is compressed and material, possibly with a different composition or texture, is excavated from shallow depths. If the higher normal albedo of the rim material indicates a compositional or textural difference from intergroove materials, then this is more consistent with the 'fracture and degassing' and 'ejecta re-impact' models of groove origin. The 'fracture and drainage' model can explain the higher normal albedo of groove rims only by the lesser maturation of the regolith.

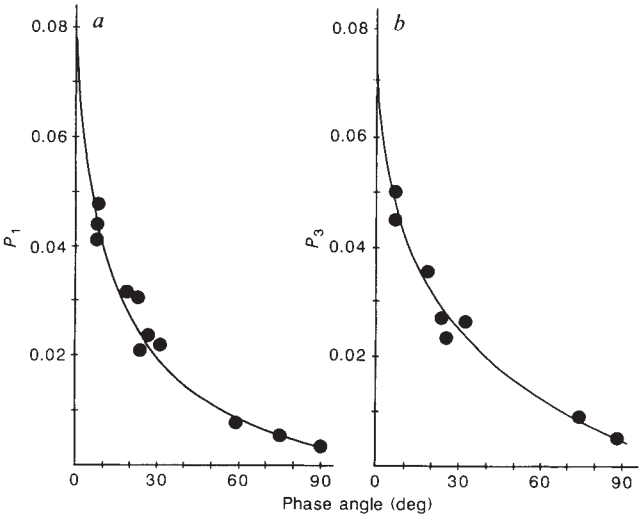


FIG. 3 The disk-integrated reflectance (P) of Phobos against phase angle at visible (channel 1, *a*) and near infrared (channel 3, *b*) wavelengths. The ground-based calibrations of the VSK absolute sensitivity were used, and the correction of CCD sensitivity for temperature was applied. The error of relative brightness measurements is 5% in both channels, and the uncertainty of absolute calibration is 10% in channel 1 and 15% in channel 3. Changes in the visible area of Phobos were accounted for by assuming that Phobos is a triaxial ellipsoid $26.6\times22.2\times18.6\,km^2$[16]. The experimental points were fitted to the model phase function[14], with the geometrical albedo p and the slope parameter Γ being free parameters. Their best fitting values are: $p=0.080\pm0.010$, $\Gamma=0.046\pm0.035$ (visible) and $p=0.071\pm0.013$, $\Gamma=0.252\pm0.085$ (near infrared). Error bars for geometrical albedo include both statistical and absolute calibration errors. Earlier values of $p=0.07\pm0.01$ and $\Gamma=0.044$ had been found for Phobos at visible wavelengths[14]. Visible and near-infrared images were taken shortly after one another, with subspacecraft latitude being constant to within 0.01° and subspacecraft longitude changing by less than 2°. The scatter of points is correlated in both channels and may be partly accounted for by deviations of Phobos's figure from an ellipse.

The VSK data obtained in the range of phase angles 7–90° supplement the ground-based[11] and spacecraft[12,13] measurements of the disk-integrated reflectance of Phobos at visible wavelengths and allows comparison of these measurements with near-infrared data (Fig. 3). At visible wavelengths the new data agree excellently with the previously determined phase function[13,14], which approximates all previous ground-based and spacecraft measurements. The geometrical albedo of 0.080 ± 0.010 is slightly larger, but is consistent, to within the stated accuracy, with the previous value of 0.07 ± 0.01. At near-infrared wavelengths the geometrical albedo of 0.072 ± 0.013 does not differ significantly from the geometrical albedo at visible wavelengths, but the phase function curve is found to be shallower (the slope parameter is distinctly different at least at the 1σ level). Either the surface of Phobos exhibits different photometric behaviours at different wavelengths or, as continuing research suggests, this effect may result from areas imaged at higher phase angles being slightly redder in colour than areas seen at low phase angles.

An important auxiliary task of the TV imaging of Phobos was to update the orbital elements of Phobos to aid the spacecraft manoeuvres on the approach to Phobos. The error in predicting the location of Phobos was decreased by an order of magnitude to several kilometres. The navigation tracking of the spacecraft resulted in a more accurate estimation of the Phobos mass, that is $1.08\pm0.01\times10^{19}\,g$[15]. Taking from model calculations, the volume of Phobos to be $5{,}530\pm300\,km^3$ (ref. 16), a mean density of $1.95\pm0.1\,g\,cm^{-3}$ is then obtained, lower than the value of $2.2\pm0.2\,g\,cm^{-3}$ based on Viking data[17]. This value is significantly less than the densities of materials having spectral reflectances similar to that of Phobos[18] (for example, hydrous and anhydrous carbonaceous chondrites and optically blackened mafic material[19–21]) strongly suggesting that the interior of Phobos could be more or less porous or contain ice.

Further analysis of the VSK data is now in progress, with emphasis on (1) updating the three-dimensional model of Phobos and topographic and geological maps of its surface; (2) characterizing spectral reflectance and photometric properties of the surface of Phobos, and identifying the implications of these results for surface composition and texture, and mechanisms of formation and evolution of craters and grooves; and (3) improving determination of the shape, volume and forced libration amplitude of Phobos, and searching for the offset of the centre-of-mass from the geometrical centre to place additional constraints on internal structure and mass distribution. □

Received 19 June; accepted 23 August 1989.

1. Thomas, P. *Center for Radiophysics and Space Research Rep. No. 693* (Cornell University, Ithaca, New York, 1978).
2. Thomas, P. *et al. J. geophys. Res.* **84,** 8457–8477 (1979).
3. Duxbury, T. C. & Callahan, J. D. *Icarus* **77,** 275–286 (1989).
4. Turner, R. J. *Icarus* **33,** 116–140 (1978).
5. Thomas, P. *Icarus* **40,** 223–243 (1979).
6. Veverka, J. *Vistas Astron.* **22,** 163–192 (1978).
7. Shkuratov, Yu. G. *Cinematics and Physics of Celestial Bodies* **4,** 33–39 (1988).
8. Head, J. in *Scientific and Methodological Aspects of the Phobos Study* (IKI, Moscow, USSR) 61–69 (1986).
9. Head, J. & Cintala, M. in *Reports of the Planetary Geology Program 1978–1979, NASA TM-80339,* 19–21 (1979).
10. Wilson, L. & Head, *J. Lunar planet. Sci.* **20,** 1211–121 (1989).
11. Zellner, B. N. & Capen, R. C. *Icarus* **23,** 437–444 (1974).
12. Noland, M. & Veverka, J. *Icarus* **28,** 404–414 (1976).
13. Pang, K. D. *et al. J. geophys. Res.* **88,** 2475–2484 (1983).
14. Lumme, K. in *Scientific and Methodological Aspects of the Phobos Study* (IKI, Moscow, USSR) 97–103 (1986).
15. Kolyuka, Yu. F. *et al. Soviet Astr. Lett.* (submitted).
16. Duxbury, T. C. *Icarus* **78,** 169–180 (1980).
17. Williams, D. G., Duxbury, T. C. & Hildebrand, C. E. *Lunar planet Sci.* **19,** 1274 (1988).
18. Wasson, J. *Meteorites,* 175 (Springer, New York, 1974).
19. Johnson, T. V. & Fanale, F. P. *J. geophys. Res.* **78,** 8507–8518 (1973).
20. Bibring, J.-P. *et al. Nature* **341,** 591–593 (1989).
21. Britt, D. & Pieters, C. *Astr. Vestnik* **22,** 229–239 (1988).

Spatial variations in thermal and albedo properties of the surface of Phobos

L. V. Ksanfomality*, V. I. Moroz*, J. P. Bibring†, M. Combes‡, A. Soufflot†, O. F. Ganpantzerova*, N. V. Goroshkova*, A. V. Zharkov*, G. E. Nikitin* & E. V. Petrova*

* Space Research Institute, Profsoyuznaya, 84/32, 117810, Moscow GSP-7, USSR
† Institute d'Astrophysique Spatiale-Orsay, 91406 Orsay, France
‡ Observatoire de Paris-Meudon, 92190 Paris, France

IMPORTANT information about the properties of the outermost layer of regolith on Phobos can be obtained from infrared thermal radiometry and spectrophotometry of the reflected radiation. From the analysis of Mariner 9[1,2], Viking 1,2[2–4] and ground-based experimental data it was concluded that the surface layer of Phobos consists of fine-grained material (regolith)[1] with a composition close to that of carbonaceous chondrites[2,3]. In the course of these measurements, Phobos was observed as a point-like object. Here we present results obtained with a resolution 1/20–1/40 of the mean Phobos diameter showing, for the first time, inhomogeneities in the thermal and spectral properties of the surface of Phobos.

The combined multichannel instrument KRFM on board the Phobos 2 spacecraft consisted of a radiometer with an angular resolution of 30′, capable of measuring thermal emission in the range 6–50 μm, and a photometer having an angular resolution of 15′, detecting in nine bands in the spectral range 320–600 nm. Radiometer channels 1–6 covered the ranges 4.8–6.4, 6.1–7.8, 7.2–9.3, 9.2–13.6, 14.3–15.9 and 16.5–49 μm, respectively, and spectrophotometer channels 1–9 measured radiation of wavelengths of 320, 328, 346, 363, 410, 445, 488, 550 and 600 nm. The spectral response of the photometer channels were gaussian, and the optical axes of the radiometer and the photometer were coincident.

The orbiting instrument was used to study Mars for two months; those results will be published elsewhere. On 25 March 1989, one of the last days on which Phobos 2 functioned, we obtained observations of Phobos from a height of 190 km. When the instrument was switched on, Phobos was already in the field of view. The relative spacecraft speed was 43 m s^{-1} and for

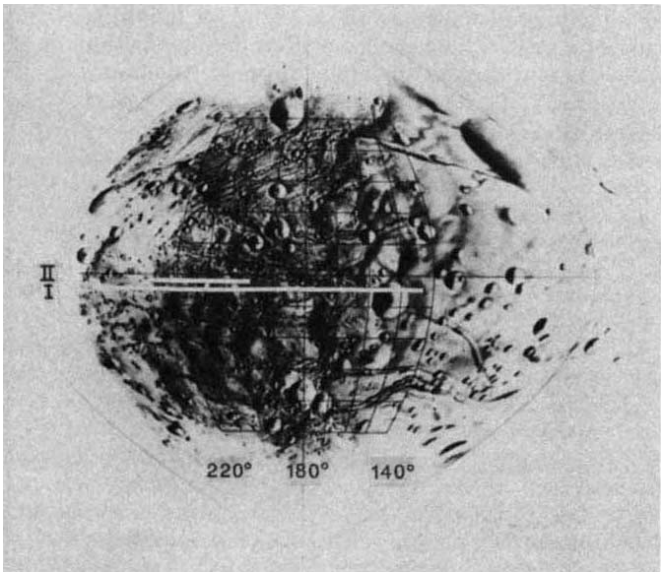


FIG. 1 The track positions on the surface of Phobos.

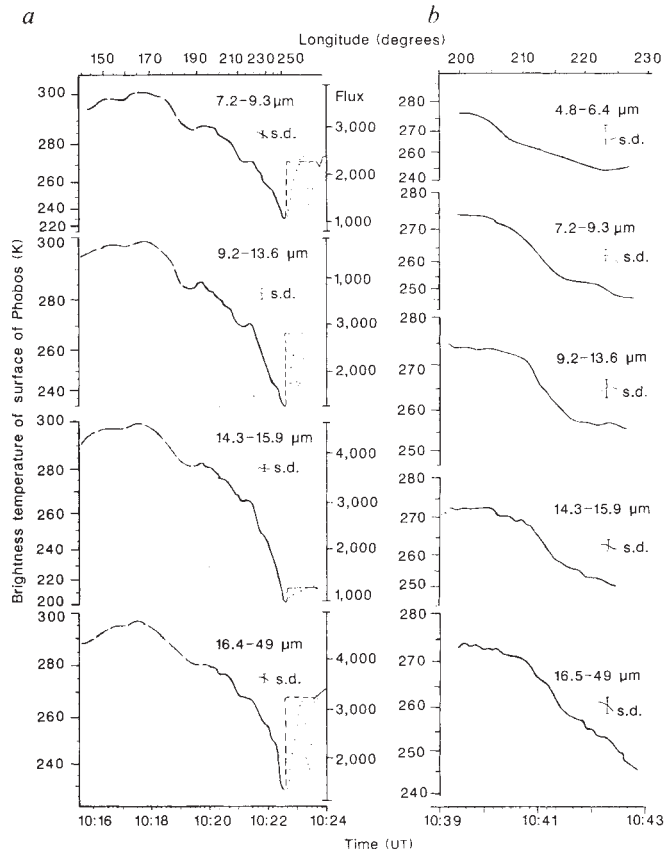


FIG. 2 The profiles of brightness temperature for *a*, track I (latitude −4.5°), radiometer channels 3–6 from top to bottom and *b*, track II (latitude −2.7°), radiometer channels 1 and 3–6 from top to bottom. Standard deviations are shown for each profile. Flux is given in arbitrary telemetry units.

~420 s the track covered a band on Phobos at a latitude of about −4.5°, from 140° longitude to the limb which was at 260° longitude (west). The surface resolution was 1.9 km and 1 km for the radiometer and photometer, respectively. After this set of observations the spacecraft turned back and measurements were obtained along a second track at −2.7° latitude, 196–229° longitude (not reaching the limb). Figure 1 shows the track positions on the surface of Phobos. The tracks crossed a number of depressions (180°–195°, 200°–210° and 215°–240°).

The profiles of brightness temperature T_B for both tracks are shown in Fig. 2. The optical axes of the radiometer channels did not strictly coincide as can be seen by comparing the profiles for 7.2–9.3 and 14.3–15.9 μm and for 9.2–13.6 and 16.5–49 μm (Fig. 2). This offset results in a difference of about half of a field of view (~1 km at the surface).

Determining surface local geometry, solar and spacecraft zenith angles, local solar time and the coordinates of the observed point, is a rather complicated problem for Phobos. The non-regular shape of Phobos cannot be described by a simple mathematical model. The normal to the model surface can be significantly different from that to the physical surface. Trajectory calculations of the time of limb crossings were inaccurate at first—the discrepancy being several minutes. Coordinates shown in Figs 2–4 and the local solar time may be in error by up to 30 s on the timescale of Fig. 2, 4 and by up to 3 min on the timescale of Fig. 3.

The highest brightness temperatures on the surface of Phobos were 296–302 K (r.m.s. ±2K), which is in good agreement with calculations[5]. These were measured in the region with the local solar time 13:15. The lowest temperature for the morning limb did not exceed 200 K.

In the first track Phobos was observed against the background

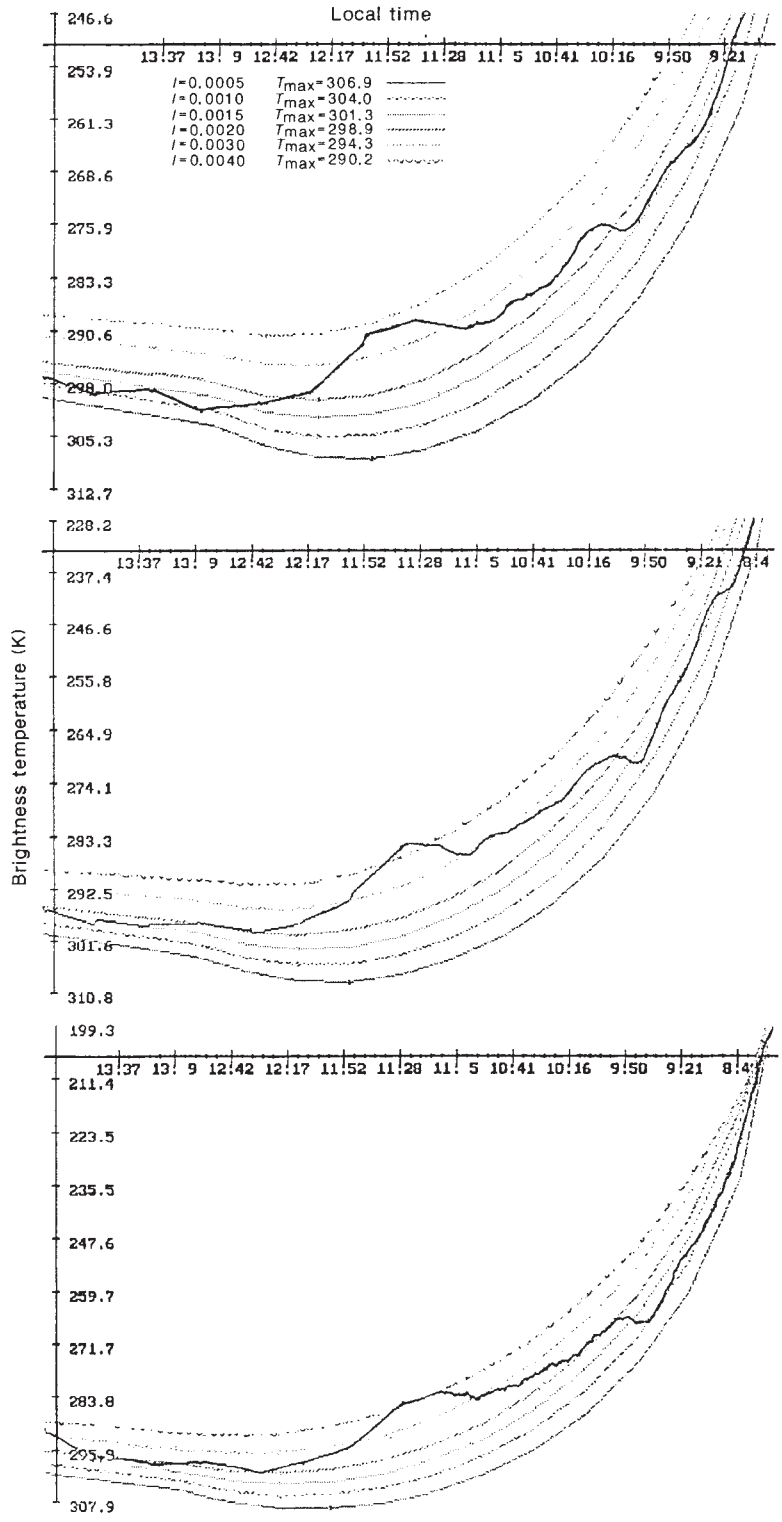


FIG. 3 The comparison of experimental curves of brightness temperature in different radiometer channels (3–5, from top to bottom covering the ranges 7.07–9.57, 9.19–13.83 and 13.70–16.72 μm, respectively) for track I with calculations. The units of thermal inertia I are cal cm^{-2} deg^{-1} s$^{-1/2}$.

of Mars and after crossing the limb the instrument viewed Mars near the local noon. The brightness temperature of Mars in this region was ~263–268 K with the exception of the 15-μm channel where it was lower ($T = 200$ K) because of the absorption of the surface emission by the martian atmosphere. The distribution of points in Fig. 2 after the departure from Phobos may be explained by the detection of radiation from Mars for a time and then from Phobos again, after the spacecraft turned back.

We compared the brightness-temperature curves along the solar local-time axes to calculations. The calculations assumed a flat semi-infinite surface layer[6], using parameters of the solar flux in the function of solar time and the surface albedo. The parameter is thermal inertia $I = (\kappa\rho c)^{1/2}$, where κ is the thermal-conductivity coefficient, ρ is the density and c is the specific heat. The calculated values are shown in Fig. 3. Channel 5 (14.3–15.9 μm) is most convenient because there is almost no distortion of the data as a result of the radiometer detecting radiation from Mars in the Phobos limb observations.

The measurements were made 37 days after the martian spring equinox. The Phobos orbit lies in the equatorial plane of Mars,

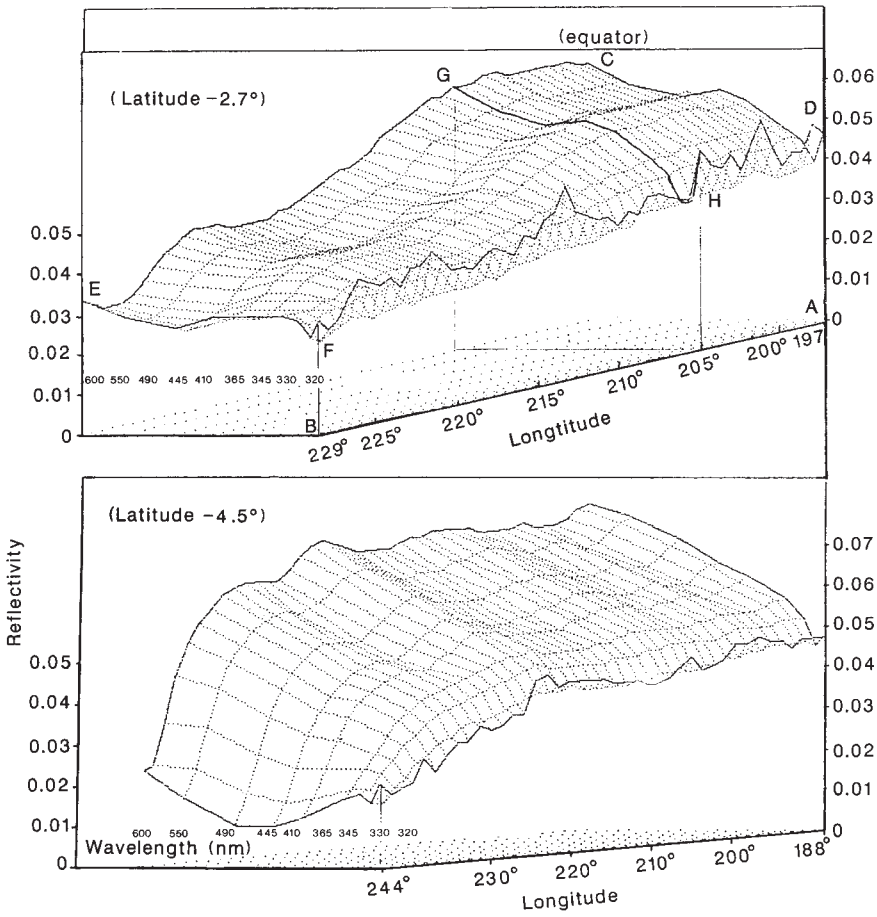


FIG. 4 The reflectivity distribution of the surface of Phobos along track II (latitude −2.7°) and along track I (latitude −4.5°) as a function of wavelength (the line G–H is for 205° longitude) and of longitude along the line A–B. As a result of the small ranges and values of angles (the phase angle was 29°, the solar zenith angle was 21.7°–50° and the spacecraft zenith angle was 3–21°) we did not correct for distortions that result from the viewing geometry.

so at local noon on Phobos the Sun was 6.5° north of the equator, and the heliocentric distance was 1.595 AU. For these conditions the calculated brightness temperature for the layer with $\kappa = 0$ was 310 K. The difference between this model and the observed temperature (~10 K at local noon) is due to the downgoing thermal flux transferred when thermal conductivity is non-zero.

The comparison (Fig. 3) shows that, except for the time interval 10:00–12:00 (longitudes indicated in Fig. 2), the thermal-inertia parameter for the surface of Phobos is close to 10^{-3} cal cm^{-2} deg^{-1} s$^{-1/2}$, which is quite close to the earlier estimates for Phobos[1,4] and the Moon[6]. For some places I was 2×10^{-3} and for the time interval 11:00–11:40 I was as high as 4×10^{-3}. The range of $I = 8\times10^{-4}$–2×10^{-3}, corresponds to fines in the vacuum having mean grain sizes tens to hundreds of micrometres[7], or a higher abundance of coarse material. Even greater values for the 11:00–11:40 time interval could be the result of localized exposures of consolidated rocks or because the layer of fines covering them is very thin. Here the zenith-angle cosine for the Sun is only slightly (7–8%) less than the maximum. The calculations show that the uppermost layer of regolith with the same properties should be only 6 K cooler, not 18 K. The albedo data in Fig. 4 do not show higher reflectivity for this region so the variation of the absorbed solar radiation could not cause this difference.

Because of the short rotation period of Phobos (7 h 39 min 14 s) the depth l_τ of thermal-wave penetration into the Phobos soil is small compared with that of the Moon. For regions with $I = 10^{-3}$, $l_\tau = 0.6$ cm and for $I = 4\times10^{-3}$, $l_\tau \approx 1.4$ cm.

In the upper and middle panels of Fig. 3 measurements obtained in wavelength ranges 7.2–9.3 and 9.2–13.6 μm are plotted. Because of the high martian temperature at these wavelengths it is difficult to trace the temperature curve at the limb of Phobos. Otherwise, they are of the same shape as for the 15-μm channel.

The photometric properties of the surface of Phobos were also found to be very inhomogeneous. Figure 4 shows the reflectivity-factor distribution of Phobos soil along each of the two tracks for nine spectral bands in the range 320–600 nm. The angular resolution of 15′ produced a spatial resolution of 900–920 m on the surface of Phobos. The field in Fig. 4 is placed in the same way as in Fig. 1 and corresponds to the spacecraft movement along line A–B from right to left.

Along almost the entire length of track I and in the first half of track II the reflectivity increases moderately in the wavelength range 320–600 nm. There are two noticeable decreases in reflectivity along track II at 450 nm and in a rather narrow band at 330 nm. Changes in reflectivity along the track II in 320-nm band (the upper panel of Fig. 4) may be caused by the noise in this channel, the cutoff in spectral sensitivity and/or high amplification. The number of points averaged together (four points for each line of spectra) is the same for both tracks. For all other channels the signal-to-noise ratio was $>10^3$.

The most interesting feature of the track II data is the wide depression: the absorption band between 400 and 600 nm in the longitude range 215°–224°. The same region shows a noticeable decrease in the brightness temperature (Fig. 2). Thus inhomogeneity of the Phobos surface properties is seen in the reflectivity data as well as in the thermal features. Although the identification of the surface composition by reflectivity is difficult, the analysis of combined data (our data and the measurements in the wavelength range 0.8–3.2 μm (the infrared imaging spectrometer experiment[8]) and spectrometric data obtained with the television system[9]) could allow the variations of reflectivity to be related to surface soil composition. □

Received 23 June; accepted 7 August 1989.

1. Gatley, I., Kieffer, H., Miner, E. & Neugebauer, G. *Astrophys. J.* **190,** 497-504 (1974).
2. Pang, K. D. *et al. Science* **199,** 64-66 (1978).
3. Pollack, J. B. *et al. Science* **199,** 66-69 (1978).
4. Lunine, J. I., Neugebauer, G. & Jekosky, B. M. *J. geophys. Res.* **87,** 10297-10305 (1982).
5. Kührt, E. *Icarus* (submitted).
6. Wesselink, A. F. *Bull. Astr. Inst. Neth.* **10,** 351-360 (1948).
7. Wechsler, A. E. & Glaser, P. E. *Icarus* 335-352 (1965).
8. Bibring, J. P. *et al. Nature* **341,** 591-593 (1989).
9. Avanesov, G. A. *et al. Nature* **341,** 585-587 (1989).

ACKNOWLEDGEMENTS. We thank Dr T. Duxbury for the photographic maps of Phobos, Dr N. Havoncon for the ballistic calculation and Dr E. Kührt for useful remarks.

Results from the ISM experiment

J.-P. Bibring*, M. Combes†, Y. Langevin*, A. Soufflot*, C. Cara†, P. Drossart†, Th. Encrenaz†, S. Erard, O. Forni*, B. Gondet*, L. Ksanfomality‡, E. Lellouch†, Ph. Masson§, V. Moroz‡, F. Rocard*, J. Rosenqvist† & C. Sotin§

* Institut d'Astrophysique Spatiale, 91405 Orsay, France
† Departement de Recherche Spatiale, Observatoire de Paris, 92125, Meudon, France
‡ Space Research Institute, Profsayuznaya 84/32, 117810 Moscow, USSR
§ Laboratoire de Geodynamique Interne, 91405 Orsay, France

THE imaging spectrometer ISM, operating in the near infrared, is the first ever flown in space for planetary observation. With a spectral range of 0.76–3.14 μm, its main goals were to obtain information on the mineralogical composition of the surfaces of Mars and Phobos, and on the composition, spatial and temporal variations of the martian atmosphere. The infrared imaging spectrometer ISM acquired 40,000 spectra of Mars and Phobos. Here we present the first results on the surface mineralogy of Mars and Phobos, and the atmospheric composition of Mars.

The first (1.65-3.16 μm) and second (0.76-1.54 μm) orders of the grating spectrometer are separately focused onto two 64-pixel PbS arrays by a dichroic beam-splitter. The resulting resolution is 0.025 and 0.0125 μm per pixel in the first and second order, respectively. The telescope is 2.5 cm in diameter. Imaging is obtained by combining the drift of the spacecraft with the scanning motion of a mirror, which can be moved by up to 20° off the optical axis. The instrument field of view (IFOV) is 12 arcmin. The detectors are cooled down to a temperature of ~200 K by a passive radiator. The sensitivity of the instrument is very high—when observing Mars, the signal-to-noise ratio is > 1,000, at 1.8 μm, for an integration time of 0.5 s.

For each resolved area of Mars, or 'spatial pixel', ISM gives an infrared spectrum composed of the 128 data extracted from the PbS detectors. During one sequence of observation, the region mapped has a width that depends on the selected scanning range (1-25 pixels) and a length depending on the duration of the sequence. Here, we refer mostly to the results obtained when the spacecraft was three-axis stabilized. ISM has observed Mars twice from a 'low' altitude (≤ 2,000 km). For these 'high-resolution' tracks, $20 \times 1{,}650\ km^2$, the pixel size is $5 \times 5\ km^2$. The regions mapped extend over the Tharsis Montes, through Pavonis Mons for one track, and through Biblis and Ulysses Paterae for the other track. From the circular orbit, at an attitude of 6,300 km, ISM mapped nine regions with a pixel size of $\sim 20 \times 30\ km^2$. Seven of these regions are located in the western hemisphere, at latitudes lower than 20° and longitudes ranging from 20° to 180°; they map more than 25% of the Tharsis Montes and Valles Marineris area. Two additional regions are located in the eastern hemisphere, within the old cratered terrains south of Arabia Terra for one, Isidis Planitia and Syrtis Major for the other. The total number of spectra obtained exceeds 36,000. They sample most of the major geological formations of Mars, with the exception of the polar caps. The heliocentric longitude of the Sun increased from −4° to 32°, relative to the spring equinox. On 25 March, ISM observed Phobos twice, when the spacecraft was tilted to point towards Phobos from a distance of ~200 km; the resolution achieved by ISM was ~0.7 km per pixel. ISM acquired about 600 spectra: one track of 1×200 pixels, along the equator of Phobos, and a region 20×20 pixels in size, mapping about one-third of the lighted hemisphere.

Below 3.2 μm, the infrared spectrum of Mars originates almost entirely from reflected sunlight. Intense bands are due to absorption by atmospheric CO_2, the strongest being centred at 2.75 μm (Fig. 3). However, it cannot easily be used to monitor the total amount of CO_2 in the atmosphere, because of saturation. Scattering is clearly seen in spectra obtained near the limb. However, scattering does not prevent observation of the surface, even at very oblique thicknesses. All observations were performed at low phase angle. For atmospheric studies, we have selected spectra obtained at a low angle of incidence (<15°). In view of current models of martian hazes, the contribution of scattering in the spectra is expected to be low, especially in the high-altitude range (10-20 km) used for the altimetry measurements described below. Consequently, we have used the 2.0-μm band to evaluate the global CO_2 pressure for each resolved spatial pixel, and derive its altimetry. We made the calculations with a band model in the effective-pressure approximation, using the formulation of Wallace *et al.*[2]. The parameters of the model are the surface pressure, the near-surface temperature and the airmass factor. To achieve the highest accuracy, we analysed 'ratioed' spectra (calculated by dividing spectra obtained at different altitudes) to eliminate instrumental effects as well as the average mineralogical signatures, and we have compared them with the ratios

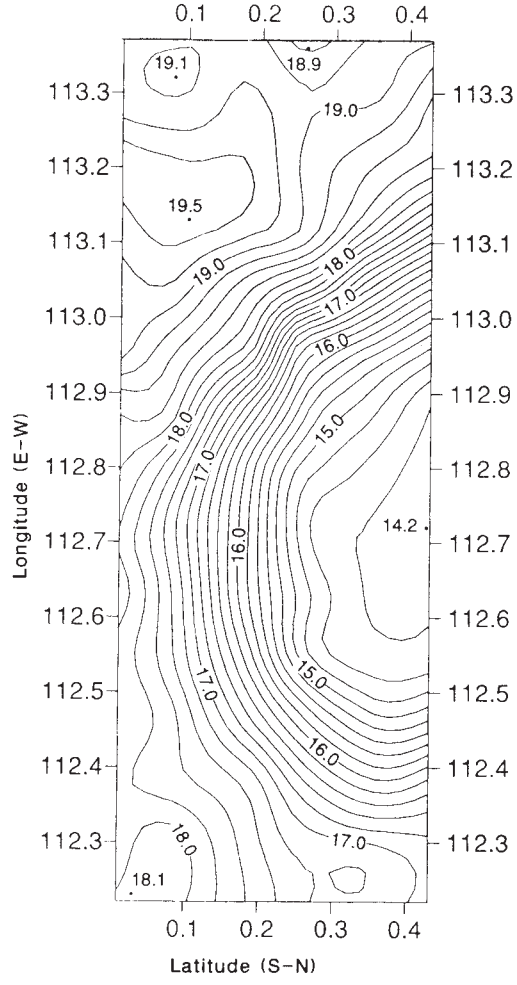


FIG. 1 Altimetry of the caldera of Pavonis Mons, derived from the CO_2 ground pressure. The contour lines are 200-m apart.

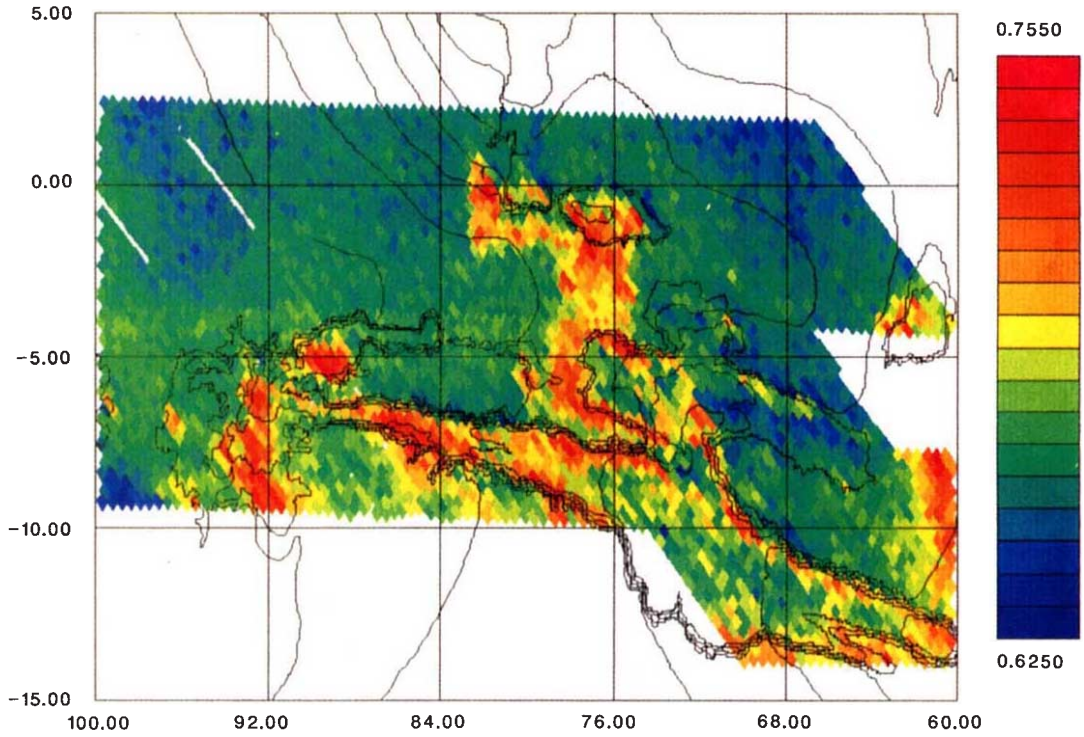


FIG. 2 Variations of the hydration feature in the Valles Marineris region, obtained on 7 March (central part), 12 March (lower right) and 26 March (upper part). The red to blue scale corresponds to an absorption of 25% to 40% at 3 μm.

of synthetic spectra. Figure 1 illustrates the altimetry data derived from a high-resolution track; the caldera of the Pavonis Mons is resolved with a pixel size of ~5 km. We have built a reference spectrum by averaging 25 spectra around the summit. We then chose 90 spectra on a regular grid covering the caldera, and divided them by the reference spectrum. Measuring the altitude at a given point from ratioed spectra requires knowledge of the vertical thermal profile at this point, as well as the altitude and pressure of one reference point. In this preliminary calculation, we assumed the atmospheric temperature at a given altitude to be spatially constant. The thermal profile has been estimated from the Viking IRTM and lander measurements[3,4], with an atmospheric temperature at ground level of 250 K, and a gradient of $-2.0\ \mathrm{K\ km^{-1}}$. The reference altitudes have been estimated from the Viking photogrammetric measurements, with the zero level of the Mariner 9 maps[5]. The altitude of the reference spectrum, near the summit, has thus been taken as 19,000 m. We obtain a depth for the caldera of 5,000 ± 300 m, with respect to the south-west rim. The high signal-to-noise ratio leads to the detection of local topographic features with heights of only 100 m. The determination of absolute altitudes is limited mainly by the reference data derived from Mariner 9, and can be provided with a precision of 5%.

ISM spectra also exhibit the signatures of CO, at 2.35 μm, and H_2O, which was studied from its bands centred at 1.84 and 2.54 μm. To evaluate the amount of these minor species, we selected 15 averaged spectra (each summing 10–30 nearby spectra) at various altitudes on Olympus Mons, from 2 to 26 km.

Ten more spectra were selected in Valles Marineris. For these averaged spectra, the spatial resolution ranges from 50 to 100 km. Assuming a column density of 2.69×10^{19} molecules $\mathrm{cm^{-2}}$ of CO above the summit of Olympus Mons (a value consistent with previous millimetre and infrared observations[5]), and no H_2O at this altitude, we obtained the following results: for CO, a column density of 7.13×10^{19} molecules $\mathrm{cm^{-2}}$ is measured at an altitude of 2 km, corresponding to a ratio for CO/CO_2 of 4×10^{-4}. For H_2O, the ratio of H_2O/CO_2 is 5×10^{-5}, with a factor of 2 uncertainty at different altitudes on Olympus Mons and Valles Marineris. This ratio corresponds to 4 μm of precipitable water (pr-μm), and is compatible with the dry limit of the water measurements of Davies[7], which range from 2 to 60 pr-μm.

The surface of Mars shows large variations in albedo and redness, at a scale corresponding to the resolution of ISM. The average surface reflectance, derived from the reflected flux at 1.1, 1.29, 1.76 and 2.45 μm, ranges from 12% to >40%, for regions observed in the same viewing conditions. These results are consistent with ground-based observations[8], with an increased dynamic range resulting from the improved spatial resolution. The redness of the continuum was derived from the ratio of the fluxes measured at 2.45 and 1.76 μm. As a general trend, bright regions are redder than darker ones. However, marked departures from this correlation are observed, particularly in the region of Valles Marineris.

The ISM spectra of Mars exhibit a broad and intense absorption feature from 2.7 to 3.14 μm. We attribute this feature to the presence of hydrated minerals. To evaluate its strength, we have divided the flux at 3.0 μm (where the CO_2 band is negligible) by that at 2.45 μm, in the continuum. We have then corrected this ratio for the redness. On average, the hydration feature is strong (35% in absorption at 3.0 μm), and exhibits large variations, by up to 20%, on the scale of the spatial resolution of ISM. Figure 2 shows a map of the strength of the hydration

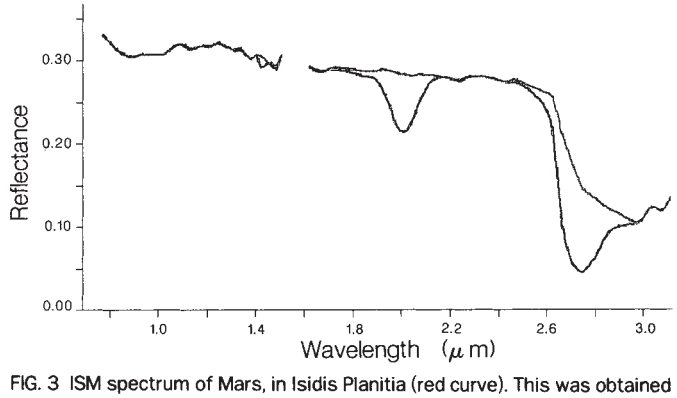


FIG. 3 ISM spectrum of Mars, in Isidis Planitia (red curve). This was obtained by dividing the raw spectrum by the solar spectrum and the instrumental transfer function, deduced from calibrations. After subtraction of features resulting from atmospheric CO_2, the spectrum (green curve) is representative of the martian surface. In particular, the broad band from 2.7 to 3.14 μm is attributed to hydrated minerals.

feature in the Valles Marineris region. When we map the Tharsis Montes volcanoes for their content in hydrated minerals, we also obtain large variations of up to 20%. In particular, when observed at high angles of incidence (50–60°), the slopes of the volcanoes appear systematically enriched in hydrated minerals relative to the much dryer surrounding plateau.

An absorption feature at 1 μm, which probably originates from Fe^{2+}-rich silicates[9], appears in all ISM spectra. Variations in this band, although small in general, are observed in most mapped regions. Processing the entire set of calibration data will, however, enable us to derive an instrumental transfer function for the 0.8–1.2-μm range, with sufficient accuracy to attribute unambiguously a given spectrum to the silicate composition of the surface.

The preliminary analysis of the two data sets on Phobos shows that the albedo of Phobos is about four times lower than that of the Tharsis Plateau surrounding Pavonis Mons, and that the hydration signature is much weaker on Phobos than on Mars. Both the very low albedo and hydration indicate a relationship with the most evolved carbonaceous chondrites (C4), rather than with the water-rich ones (C1 and C2). On a kilometre scale, the hydration and silicate features exhibit variations of up to 10%.

Although the Phobos 2 mission ended prematurely, ISM acquired a large number of spectra of Mars, sampling most of the major geological formations. The high sensitivity of ISM should provide very valuable data on the atmosphere and surface of Mars, with a resolution of a few tens of kilometres. The ISM observation of Phobos constitutes the first spectral imaging of a small body. Heterogeneities are observed on a kilometre scale, which should yield insights into the origin and evolution of Phobos. □

Received 6 June; accepted 17 August 1989.

1. Jaquin, F., Gierasch, P. & Kahn, R. *Icarus* **68**, 442–461 (1986).
2. Wallace, L., Prather, M. & Belton, M. J. S. *Astrophys. J.* **193**, 481–493 (1974).
3. Martin, T. Z. *Icarus* **45**, 427–446 (1981).
4. Seiff, A. *Adv. Space Res.* **2**, 3–17 (1982).
5. Wu, S. C. *J. geophys. Res.* **84**, 7955–7959 (1979).
6. Clancy, R. T., Muhleman, D. O. & Jakovsky, B. M. *Icarus* **55**, 282–290 (1983).
7. Davies, D. W. *Icarus* **45**, 398–414 (1981).
8. Bell, J. F. & McCord, T. B. *Icarus* (in the press).
9. Hunt, B. R. & Salisbury, J. W. *Mod. Geol.* **1**, 283–300 (1970).

Thermal imaging of the surface of Mars

A. S. Selivanov, M. K. Naraeva, A. S. Panfilov, Yu. M. Gektin, V. D. Kharlamov, A. V. Romanov, D. A. Fomin & Ya. Ya. Miroshnichenko

Space Research Institute, Profsoyuznaya 84/32, 117810 Moscow, GSP-7, USSR

THE aim of the Thermoscan experiment on the Phobos spacecraft is the further study of the thermal properties of the surface of Mars that was begun by the Soviet Mars spacecraft and the Mariner and Viking spacecrafts. These previous investigations obtained data concerning the temperature distribution in different areographic regions, the evaluation of daily and seasonal changes of the planet temperature, and the thermal inertia of the martian surface. Previous data had a spatial resolution of a few tens or a few hundreds kilometres; in particular, the resolution of the data collected by Vikings 1 and 2 is 8–175 km. The Thermoscan experiment provides thermal images of far higher resolution, giving more comprehensive data on the temperature distribution and thermal inertia of the surface of the planet and its relief elements and making it possible to search for small thermal anomalies.

Multispectral scanning is the best technique for the physical interpretation of the data obtained, providing the simultaneous detection of the brightness temperature of the sighted region and its reflectance properties, which determine the heating of the surface by the Sun's radiation.

The Thermoscan radiometer operates in two spectral bands: 8.5–12 μm and 0.5–0.95 μm. The line-by-line scanning is achieved by means of a mirror oscillating in a stepwise fashion, perpendicular to the spacecraft motion velocity; the frame scanning is accounted for by the motion of the spacecraft itself.

The instantaneous field-of-view value in both channels is 0.9 arcmin, the scanning angle 6.1° and the scanning rate 1 line per s. It provides a resolution at nadir of 1.8 × 1.8 km and a swath length of 650 km from the circular observational orbit of 6,300 km. The area traversed in the direction of motion is a function of the length of the session. To provide the radiometric measurements during the scanning process, the instrument is calibrated relative to inner reference radiators and space. The fore optics are included in the calibration with the exception of the scanning mirror. The Thermoscan thermal response characteristic is shown in Fig. 5. The first session scanning the surface of Mars was conducted on 11 February, 1989 from the intermediate elliptical orbit, which was the transfer stage to the circular orbit. This orbit provided a unique opportunity to observe from altitudes lower than that of the circular orbit, (in the range from 1,150 to 5,200 km). From these altitudes, a surface resolution of 300 m was obtained. Figure 1*a* presents the scanning geometry in this session.

The scanning was conducted under conditions of uninterrupted astro-solar spatial attitude control of the spacecraft. Moreover, the axis-of-sight position in the centre of Thermoscan line is aligned with the solar-ray direction. During scanning, the distance from the spacecraft to the planet surface and the velocity of the subsatellite point were varied. As a result, changes in

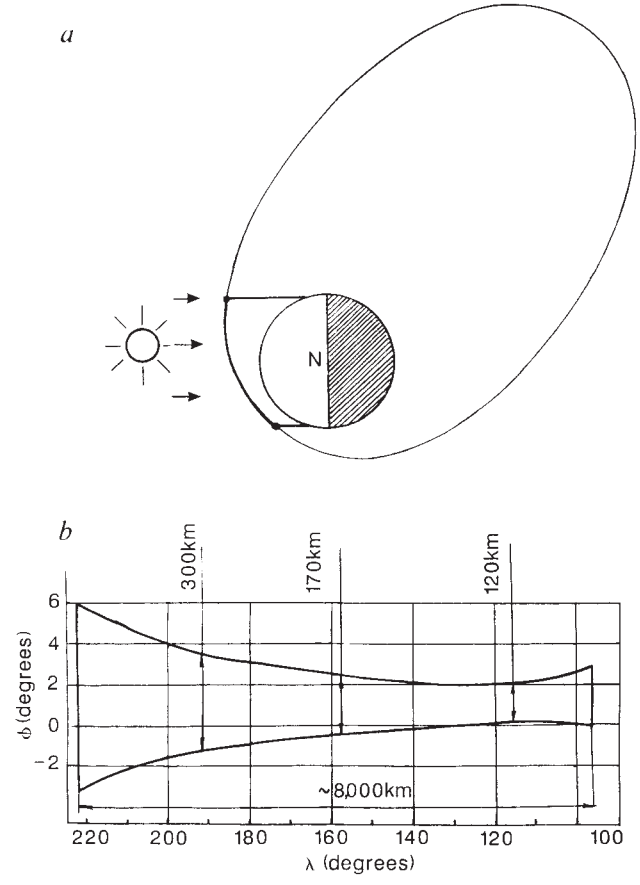


FIG. 1 Scanning geometry in the session of 11 February 1989. *a*, Position of the spacecraft in scanning session. *b*, The areographic coordinates of the scanned area.

dimensions of the area traversed in the line direction and in line and frame resolution occurred.

Temperature profiles on the surface with gaps between them were obtained. The gaps have variable sizes due to variations in subsatellite point motion velocity. The image synthesis here, therefore, causes a number of significant geometrical distortions to appear. Nevertheless, after completion of the partial geometrical correction, images of high quality have been obtained. Figure 2*a* shows the image fragments: 8.5–12 μm; 0.60–0.95 μm. The areographic coordinates (λ-longitude, φ-latitude) and local time of the scanned area are presented in Fig. 1*b*.

The zenith angle of the Sun during the scanning process varied from 10° to 90° and then to 20°. The seasonal conditions of the scanning experiment correspond to the astronomical martian spring (areocentric longitude of Sun, $L_S = 357°$). The scanning was conducted in near-equatorial, light and flat areas (the regions Elysium Aeolis, Amazonis, Memnonia, Tharsis), characterized by the existence of fractures of relief and relatively few craters of meterorite origin. At the end of the session, one of the highest martian mountains, Pavonis Mons, with the volcanic crater on the peak, fell within the instrument's field of view.

Figure 3 illustrates the scanning geometry for the circular orbit (on 1 March 1989 at an altitude of 6,300 km). The line and frame scanning are matched, and the geometrical distortions during the scanning accompanied by the continuous astro-solar control of the spacecraft spatial attitude, arises only because of planet sphericity. Because of this condition and as a result of intricate terrain relief and the marked inhomogeneity of the reflectance characteristics in the scanning area, images received in the second session contained more data. Fragments of these images are shown on Fig. 2*b*. The areographic position and local time of the scanned Mars area is shown in Fig. 3*b*. The zenith angle of the Sun during the scanning varied over a narrower range than in the first session (from 55° to 90° to 30°). Seasonal conditions were the same as in the first session ($L_S = 6°$).

The scanning area covers the near-equatorial zone with a host of craters (regions Oxia Pulus, Mapgapitifer Sinus, Apabia, Sinus Sabarus) and with multiple relief forms: such as erosive form, fractures twisting valleys and so on.

Beside these distinctions, the images obtained during two sessions have much in common with each other. The contrast of the thermal images on the whole scanning route is higher than in the short-wave range because of the lower dependence of the thermal contrasts on the Sun's position during scanning. A no-shadow condition in the visible spectrum gives rise to diminished contrast in channel 0.6–0.85 μm and allows the surface reflectance to be determined in that spectral range.

Examination of the images in both channels, obtained in both sessions, shows diverse examples of direct and inverse correlations of the brightness temperature with reflectance of the surface areas; that is, the darker regions are hotter and vice versa

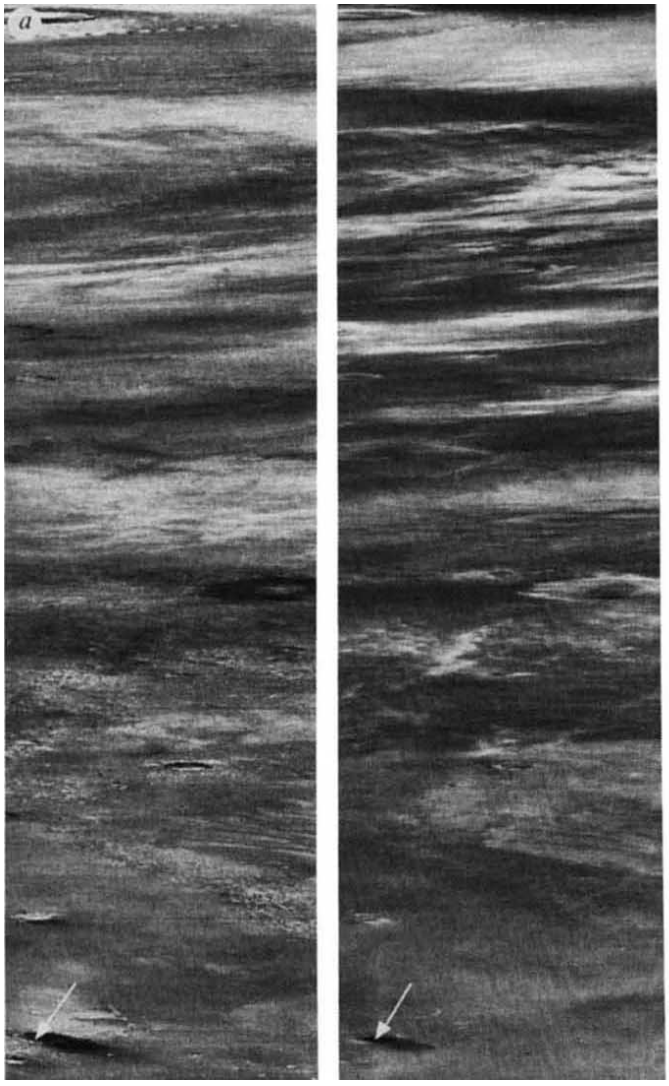
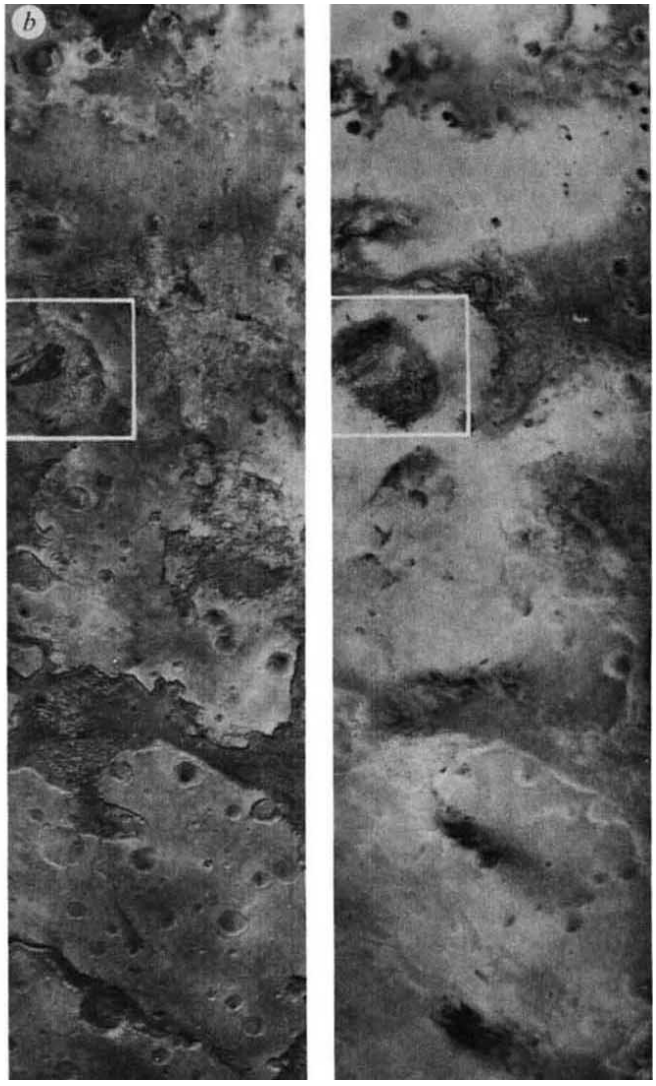


FIG. 2 *a*, The Mars surface-image fragments in the session of 11 February, 1989. Left, spectral band 8.5–12 μm; Right, spectral band 0.6–0.95 μm. *b*,



The Mars surface image fragments in the session of 1 March, 1989. Left, spectral band 8.5–12 μm; Right, spectral band 0.6–0.95 μm.

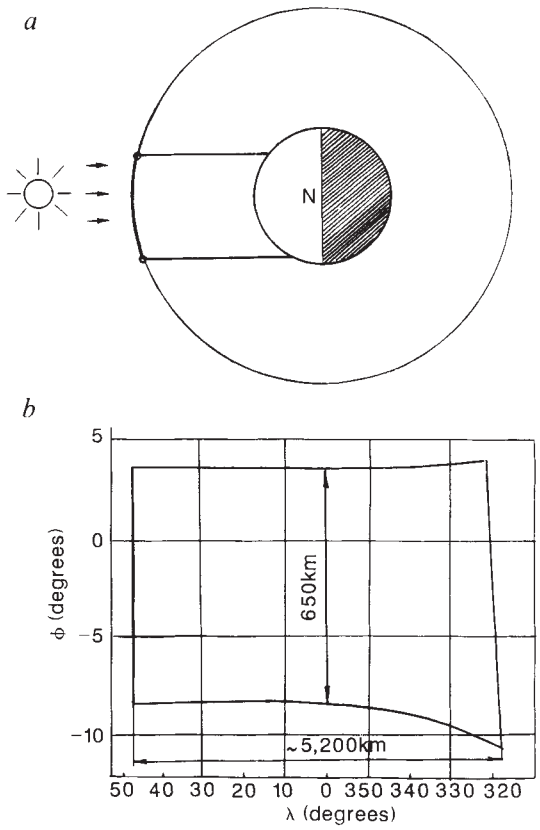


FIG. 3 Scanning geometry in the session of 1 March, 1989. *a*, Position of the spacecraft in the scanning session. *b*, The areographic coordinates of the scanned area.

(contrast inversion). This is probably caused by several factors such as thermal properties of the soil, surface-layer structure, and relief geometry. We now consider some examples of selected areas in the image fragments. The crater interior, marked by the arrow on Fig. 2*a*, is darker in the visible spectrum and, colder in comparison with the surrounding surface. Such a combination of signals could stem from the relief geometry or thermoinertial properties of the soil. The local dark surface region, lying southwest of this crater and having a lower brightness temperature, is of great interest. It can be said with confidence that it is an area with increased thermal inertia[8]. The observed temperature differences are equal to 20–25 K. Unlike this inverse correlation of temperature/brightness, the depression zone and its

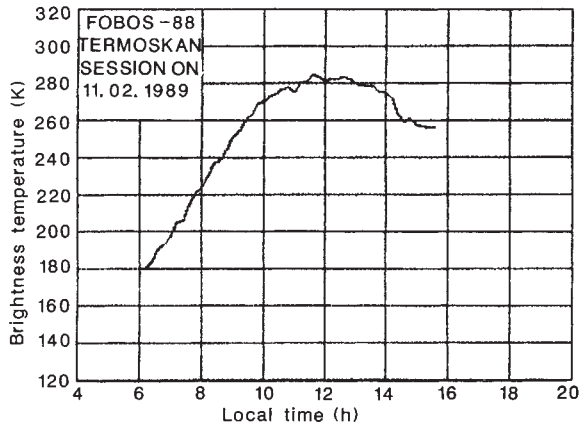


FIG. 4 Diurnal variations of the brightness temperature in the session of 11 February 1989.

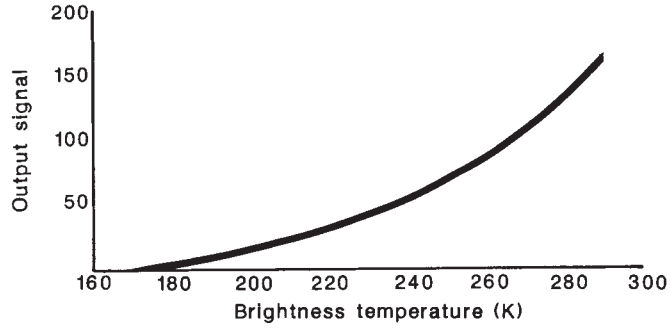


FIG. 5 Thermal response characteristic of Thermoscan.

peripheral portion (enclosed in frame on Fig. 2*b*), illustrate the direct correlation: the darker regions have elevated temperatures relative to the light regions (temperature difference is 10–15 K). During both scanning sessions the observed brightness temperatures on the whole route range from 180 to 280 K (see Fig. 4), which agrees with the previous observations and models[1–7,9,10]. New information is acquired concerning the temperature differences of separate, closely situated terrain details which reach 20–30 K.

The data presented here should be regarded as preliminary. The rich experimental information, newly gained thanks to the high spatial resolution of the Thermoscan device, demands further analysis.

After finishing the work on this article, two more sessions scanning the surface of Mars were carried out (26 March, 1989), in one of which the Phobos shadow image on the surface of Mars is visible and thermal spectral bands were obtained. The appropriate information will be given in subsequent publications. □

Received 10 June; accepted 17 August 1989.

1. Neugebauer, G., Munch, G., Kieffer, H., Chase, S. & Miner, E. *Astr. J.* **76**, 719–728 (1971).
2. Moroz, V. I. & Ksanfomality, L. V. *Icarus*, **17**, 408–422 (1972).
3. Moroz, V. I. *et al. Kosmich. Issled.* **13**, 389–402 (1975).
4. Kieffer, H. H. *et al. J. geophys. Res.* **78**, 4291–4312 (1973).
5. Ksanfomality, L. V. & Moroz, V. I. *Kosmich. Issled.* **13**, 77–79 (1975).
6. Kieffer, H. H. *et al. Science*, **193**, 781 (1976).
7. Kieffer, H. H. *et al. J. geophys. Res.* **82**, 4249–4281 (1977).
8. Christensen, P. R. *Icarus* **56**, 496–518 (1983).
9. Moroz, V. I. *Physics of Mars Planet* (Nauka, Moscow, 1978).
10. Vdovin, V. V. *Kosmich. Issled.* **15**, 238–247 (1977).

Determination of the elemental composition of martian rocks from Phobos 2

Yu. A. Surkov, V.L. Barsukov, L. P. Moskaleva, V. P. Kharyukova, S. Ye. Zaitseva, G. G. Smirnov & O. S. Manvelyan

V. I. Vernadsky Institute of Geochemistry and Analytical Chemistry, Academy of Sciences, Kosygin Street 19, 117975 Moscow, GSP-1, USSR

PREVIOUS orbiter and lander pictures have given an indication of the general morphology of the surface of Mars and of some relatively small features on the surface. Here we have determined the elemental composition of martian soil from Phobos 2 using orbital γ-ray spectrometry. This technique is used to analyse the surfaces of planets which either do not have an atmosphere or which have only a low-density atmosphere (that is, there is negligible absorption of γ-rays). The γ-radiation is caused by the decay of radioactive elements excited by the interaction of solar (SCR)

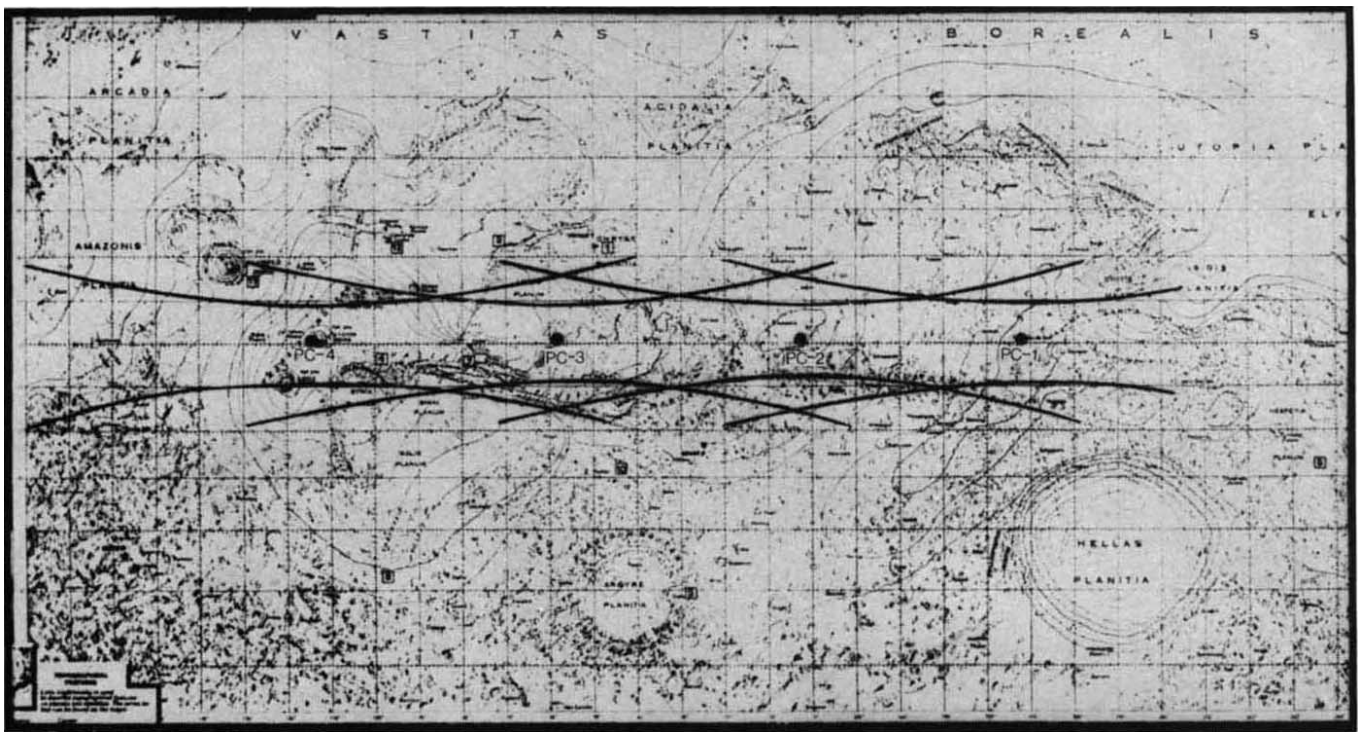


FIG. 1 Map of Mars indicating the regions where γ-radiation was measured.

and galactic (GCR) cosmic rays with the surface materials. Measurements of this radiation by a γ-ray spectrometer enable the elemental composition of the near-surface rocks to be determined. The first successful measurements of γ-radiation were conducted from Luna 10 (ref. 1). The method was used subsequently during studies of the Moon from Apollos 15 and 16 (ref. 2), and Mars 5 (ref. 3).

To measure the γ-radiation of Mars the Phobos spacecraft was provided with a γ-ray spectrometer with a 100 × 100 mm CsI(Tl) crystal. This γ-ray spectrometer was more efficient than those used earlier. The energy range was 0.3–10.0 MeV. The energy resolution of the detector (with respect to Cs^{137}) was 10%. The analyser had 512 energy channels, each with a capacity of 2×10^{16} bits. Power consumption was 13 W. The mass of the device was 12.5 kg. The device consists of three units: an outrigger detection unit, a multi-channel pulse-height analyser and an electronic unit. The optimum temperature conditions for the operation of the γ-ray spectrometer are achieved by using a heater with a thermoregulator.

In γ-ray spectrometry, the analysis of background γ-radiation sources and the reduction of their contribution to the spectrum that is being recorded are of great importance. The background radiation is caused by the decay of naturally radioactive elements that are part of the structural materials of the γ-ray spectrometer and the spacecraft, and by nuclear interactions of cosmic rays with these materials. To reduce background radiation the device must be installed as far away as possible from the main mass of structural materials of the craft. On Phobos 2, the detection unit of the γ-ray spectrometer was placed on the solar panel at a distance of 4 m from the craft's centre.

The contribution of various components of the background radiation to the spectrum being recorded differs for measurements taken close to the planet's surface and those taken at large distances from it. In measurements taken at the planet's surface, the background contribution due to the interaction of GCR with the material of the spacecraft and the γ-ray spectrometer decreases because of screening by the planet of a portion of GCR, but a flux of γ-quanta emerges. This flux is initiated by reactions of secondary neutrons with the structural materials of the spacecraft and the γ-ray spectrometer.

Background γ-radiation comes from the planet's surface and is caused by processes such as the decay of π-mesons, initiated by GCR in the planet's atmosphere and soil, γ-bremsstrahlung, the Compton effect, and so on. Such a complex background makes it impossible to exclude it in spectra measured above the martian surface, by the direct subtraction of the background measured on the Earth–Mars route. The identification of the contribution of different processes in the spectrum is required. For this purpose the detector's own background was measured in the low-background chamber at the pre-flight stage of the experiment, and statistical measurements of background spectra were repeatedly carried out on the flight route and on the sections of orbits lying far from the martian surface. These measurements have made it possible to analyse the dynamics of the change of the contribution of various processes to background radiation.

In January 1989, Phobos 2 reached Mars and was injected into orbit. For about three weeks the spacecraft was in a three-

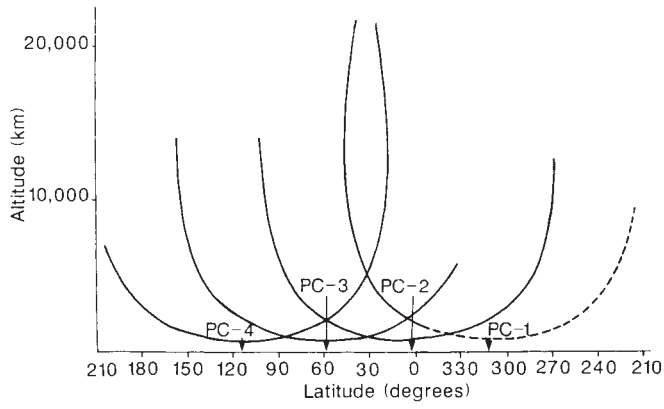
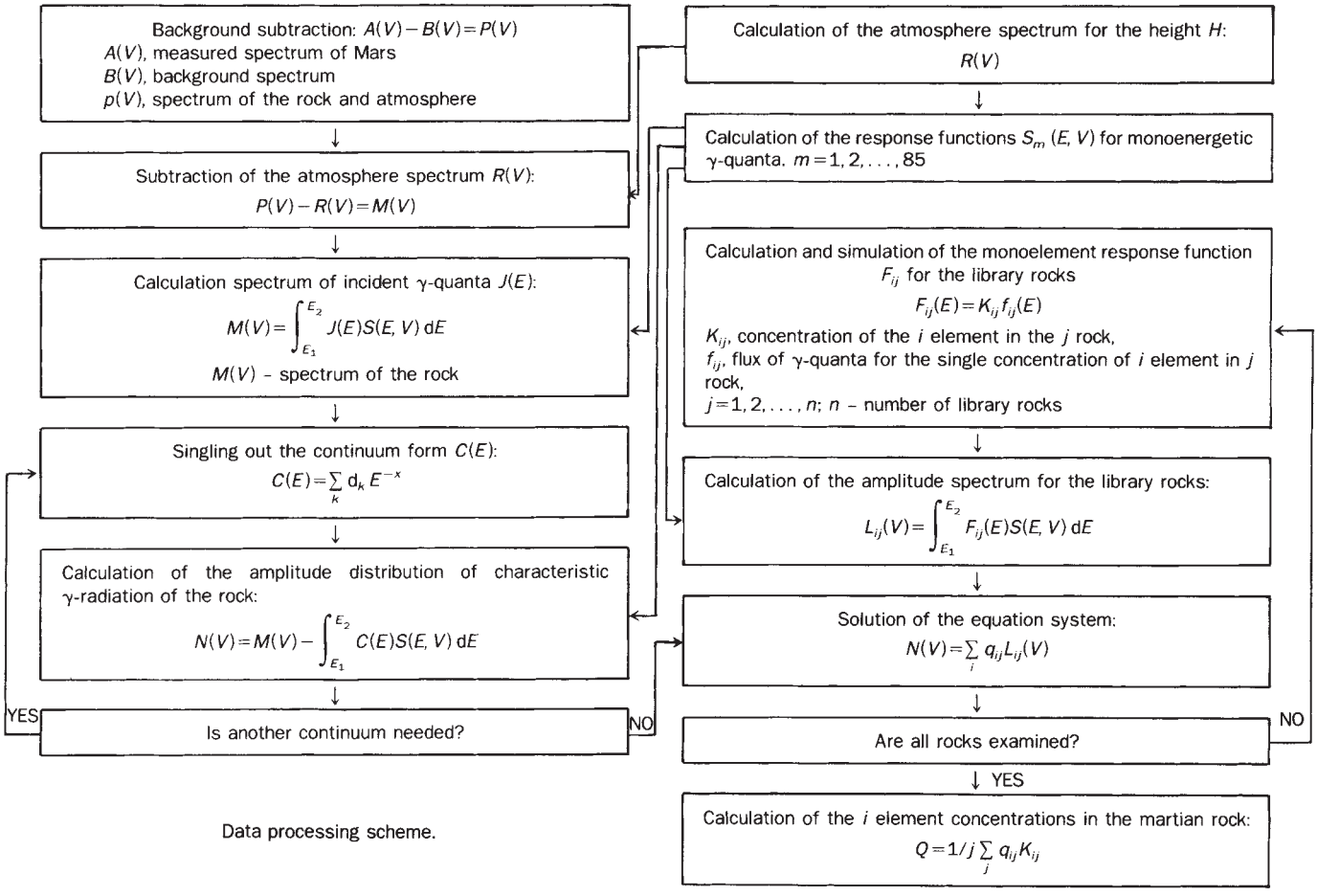


FIG. 2 Sections of the flight trajectories PC-1 to PC-4 of Phobos 2 in the vicinity of Mars on which measurements of γ-radiation were conducted.



Data processing scheme.

day elliptical orbit with a minimum and maximum distance of ~870 and ~80,000 km from the surface of the planet. The craft's orbital plane was inclined to the equatorial plane by 1°. At this stage the planet's γ-radiation was measured when the spacecraft approached the surface (in the perigee region). At the maximum distance the background radiation was measured.

Figure 1 presents a map of the martian surface. It indicates regions where the γ-radiation was measured. Points PC-1, PC-2, PC-3 and PC-4 correspond to the places where the craft was closest to the surface of Mars. Envelopes around these points are indicated. They show the limits of the planetary surface area from which 50% of the γ-radiation was collected.

Figure 2 shows sections of the craft's trajectory on which measurements were made. As can be seen from the figure, measurements of γ-radiation were started at a distance >5,000–10,000 km, then were conducted at closer distances and, finally, were completed when the craft flew away from the martian surface at a distance >5,000–10,000 km. Measurements of each γ-spectrum lasted ~10 min. Figure 3 presents a martian γ-ray spectrum measured in the PC-3 region.

Here we summarize the results of analysis of the γ-spectra measured on the lowest sections of the PC-3 orbit. The general diagram of data processing is presented in the scheme below. Analytical results are included in Fig. 3 legend; errors given are due to statistical uncertainties of measurements and uncertainties in singling out the continuum, solar and galactic ray fluxes, matrix effects that depends on the chemical composition of the rock. The errors were calculated as the r.m.s. deviation of the different solutions of the equation $N(V)=\sum_i q_{ij}L_{ij}(V)$ from the average (see block diagram).

We started processing from this particular rendezvous session because during this session the spacecraft explored the region situated close to that which was studied from Mars 5 in 1974. The elemental composition of the martian rock in the PC-3

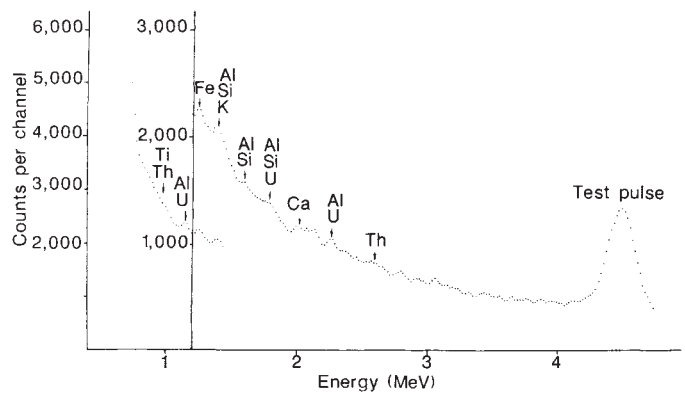


FIG. 3 Gamma-ray spectrum measured on the PC-3 trajectory at the closest distance to Mars. The measurements took 1,800 s; this time corresponds to an interval of 102°–15° latitude. The martian rock composition is:

Element	Phobos 2 (%)	Mars 5 (%)	Viking 1 (%)	Viking 2 (%)
O	48±5	44±5	50.1±4.3	50.4
Mg	6±3	—	5.0±2.5	—
Al	5±2	5±2	3.0±0.9	—
Si	19±4	14±3	20.9±2.5	20.0
S	—	—	3.1±0.5	2.6
Cl	—	—	0.7±0.3	0.6
Ti	1±0.5	—	0.51±0.2	0.61
K	0.3±0.1	0.3±0.1	0.25	0.25
Ca	6±3	—	4.0±0.8	3.6
Fe	9±3	14±4	12.7±2.0	14.2
U	$(0.5\pm0.1)\times10^{-4}$	$(0.6\pm0.1)\times10^{-4}$	—	—
Th	$(1.9\pm0.6)\times10^{-4}$	$(2.1\pm0.5)\times10^{-4}$	—	—

regions relates to a relatively large area incorporating the regions of Lunar Planum and Xanthe Terra. It is therefore hard to relate the composition to a specific geomorphological province.

For comparison, we also present the contents of the basic rock-forming elements obtained from Mars 5 (ref. 3) and Vikings 1 and 2 (ref. 4). We calculated the oxygen concentration following a similar procedure to that used for determining the concentrations of the other elements. The atmospheric oxygen content was accounted for in the calculation of the atmospheric spectrum (see block diagram).

The similarity between the rock composition in the PC-3 region and the rock composition measured from Vikings 1 and 2 allows us to conclude that vast territories on the martian surface are covered with the aeolian deposition layer.

Because Phobos 2 explored a thicker rock layer than Vikings 1 and 2 we assume that the composition derived for PC-3 reflects a mixture of surface material and the underlying bedrock. In this connection the possible compositions of the bedrock can be calculated proceeding from different contributions of both components (Yu. Surkov, personal communication). The closest analogue of the rock on Earth was determined on the basis of the systematics of magnetic terrestrial rocks, and has been found to be subalkaline basalt occurring on oceanic islands.

These, then, are the first, preliminary results of analysis of γ-ray measurements from Phobos 2. At present, data processing is in progress. It might be possible to reveal considerable differences in the composition of the bedrock lying in the different regions of the martian surface that have been explored. Of particular interest are spectra measured during the spacecraft's fourth encounter with the planet's surface, when measurements were conducted over the volcanic province of Tharsis. At the moment of closest rendezvous the craft was over the giant Pavonis volcano. Also of interest are other regions representing a great diversity of geomorphological structures on the martian surface. However, further time is required for detailed processing of the data obtained. □

Received 19 June; accepted 4 Setember 1989.

1. Vinogradov, A. P., Surkov, Yu. A., Chernov, G. M., Kirnozov, F. F., & Nazarkina, G. B. *Kosmich. Issled* **IV**, 871–879 (1966).
2. Metzger, A. E., Trombka, J. I., Reedy, R. C. & Arnold, J. R. *Proc. 5th Lunar Sci. Conf.* 1067–1078 (Pergamon, Oxford, 1974).
3. Surkov, Yu. A., Moskaleva, L. P., Manvelyan, O. S. & Kharyukova, V. P. *Kosmich. Issled.* **XVIII**, 623–631 (1980).
4. Toulmin, P. III, Clarc, B. C., Baird, A. K., Keil, K. & Rose, H. J. *Science* **194**, 81–84 (1976).

Observation of the γ-ray emission from the martian surface by the APEX experiment

C. d'Uston*, J.-L. Atteia*, C. Barat*, A. Chernenko†, V. Dolidze†, A. Dyatchkov†, E. Jourdain*, V. Khariukova‡, N. Khavenson†, A. Kozlenkov†, R. Kucherova†, I. Mitrofanov†, L. Moskaleva‡, M. Niel*, A. Pozanenko†, P. Scheglov‡, Yu. Surkov‡ & A. Vilchinskaya†

* Centre d'Etude Spatiale des Rayonnements (CESR), CNRS/UPS-BP 4346-31029 Toulouse Cedex, France
† Space Research Institute, Profsoyuznaya 84/32, Moscow 117810, GSP-7, USSR
‡ S. I. Vernadsky Institute of Geochemistry and Analytical Chemistry, Moscow, USSR

THE study of γ-ray emission from planets with a thin atmosphere allows a rough estimation of the chemical composition of their surface to be made. The first attempt to make orbital measurements of the γ-ray flux from Mars was made by the Soviet spacecraft Mars 5 (ref. 1). Here we present results from the γ-ray APEX experiment on board Phobos 2. We have measured the γ-ray albedo of the martian surface. This albedo appears as a large continuum between a few hundred keV and 8 MeV corresponding to a source intensity of 5 photons $cm^{-2}\ s^{-1}\ 2\pi^{-1}$ (estimated in the 150–1500 keV energy range). There is also evidence for a 511-keV annihilation emission and for a broad contribution of probable nuclear lines between 4 and 7 MeV.

The APEX experiment was designed primarily to record the temporal variations and the spectral characteristics of cosmic γ-ray bursts and solar flares in the 60 keV–9 MeV energy range with a fine-time resolution. However, this instrument also performed detailed γ-ray measurements when Phobos 2 was close to the martian surface. We report here preliminary results obtained near the periapsis of four successive elliptical orbits.

The APEX instrument consists of a detector built in the Soviet Union[2] with the electronics supplied by France. The detector is a 10×10 cm cylindrical CsI(Tl) crystal fixed at the edge of a

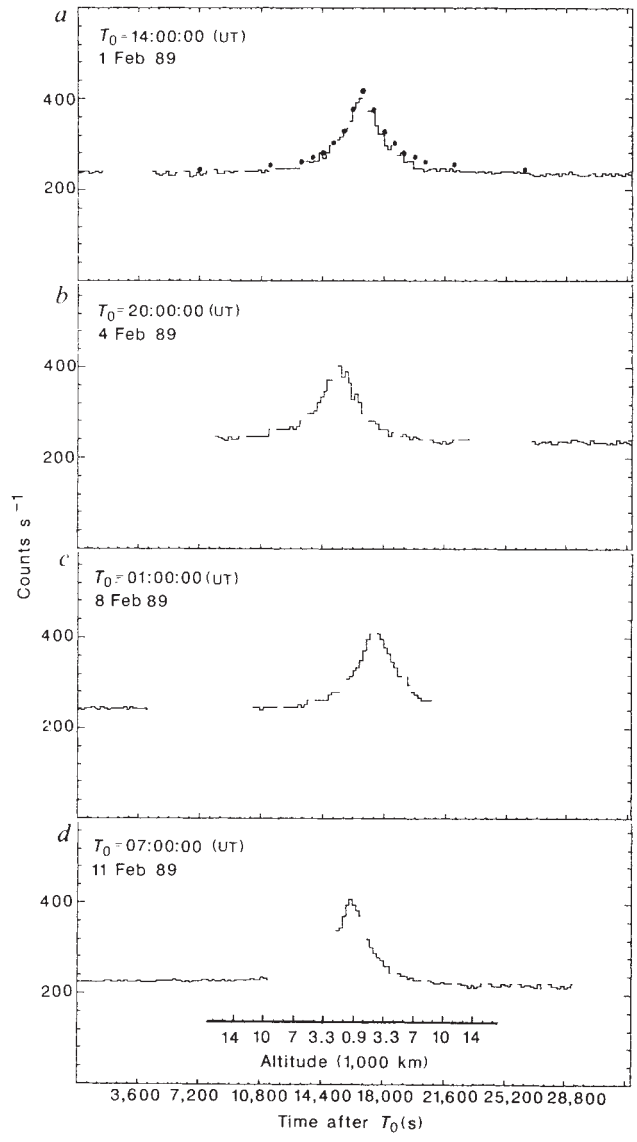


FIG. 1 Count rate through the pericentre passages in the integral channel 150–1,500 keV: *a*, *b*, *c* and *d* correspond to pericentre numbers 1, 2, 3 and 4. In the upper left corner. Time (UT) is given corresponding to the beginning of the time axis. In the top panel *a*, expected signal variation is also indicated by dots. The inset *x*-axis on the bottom panel shows the altitude above surface.

solar panel to cover a large field of view ($\sim 4\pi$ sr at >300 keV). Its relative energy resolution ($\partial E(\mathrm{FWHM})/E$, where FWHM is the full width at half maximum) is 11% at 0.662 MeV, 7.1% at 2.0 MeV, and 5.4% at 6.0 MeV. The total efficiency exceeds 0.80 in the whole energy range. The platform-mounted electronic circuit uses the capabilities of an 80C86 microprocessor to store data in coarse and fine-time resolution modes both for temporal and spectral records. The fine-time resolution mode is selected either automatically through the detection of a significant count-rate excess or by telecommands. It authorizes the recording of 116 consecutive spectra, over 108 energy channels, with a fixed number (chosen by command) of accumulated counts between 150 and 1,500 keV for each individual spectrum.

On arrival near Mars, Phobos 2 was injected into an elliptical orbit[3] with a period of 3.23 days, and altitudes above the surface at apocentre and pericentre of 79,300 and 860 km, respectively. The orbital plane was very close to the martian equatorial plane. The spacecraft followed this orbit for 4 complete revolutions before a correction of trajectory. The background count-rate was measured continuously. Close to each pericentre, the on-board software triggered 16 periods of ~9 min duration with a fine-time resolution mode for a detailed analysis of the γ-ray emission from the planet surface. This sequence was spread over 4 h centred approximately at the pericentre time. The corresponding dates, times (UT), altitudes and areographic longitudes are given in Table 1.

As shown in Fig. 1, for each pericentre passage, the count rate integrated between 150 and 1,500 keV is clearly influenced by the planet. Far from the pericentre time the total γ-ray background is steady at ~240 counts s^{-1}. At ~1.2 h before the pericentre time, corresponding to an altitude of ~8,000 km, a gradual increase of the count-rate is observed in each pericentre approach up to a maximum of ~410 counts s^{-1}. The count-rate then returns symmetrically to its steady level. This increase, by a factor 1.7, is probably due to the planetary albedo which, in γ-rays, consists mainly of a continuum produced by the scattering of the nuclear emissions generated either by the natural radioactivity or by the bombardment of cosmic-ray particles into the nuclei of the constituents of Mars.

TABLE 1 General characteristics of pericentre passages

Pericentre no	1	2	3	4
Date (1989)	1 Feb.	5 Feb.	8 Feb.	11 Feb.
Time (UT)	18:39:37	00:15:47	05:51:58	11:28:12
Altitude (km)	867	859	850	842
Longitude (deg)	310	4	59	113

As the APEX electronics does not discriminate between particles, we cannot exclude a neutron contribution. Let us assume that the count-rate increase is due to the total albedo emission; in this case, it can be verified that the temporal variation of the flux is only a geometrical effect dependent on the altitude. This is illustrated in Fig. 1 by the good agreement between the observed data obtained during the first pericentre passage and the points (large dots) deduced from this geometrical calculation, assuming that the albedo source is located above the atmosphere at an altitude of 120 km. From this dependence, it is easy to relate the measured count rate to the source intensity. If the additional count rate (~170 counts s^{-1}) at pericentre is due to the γ-ray albedo only, and using a value of ~80 cm^2 for the detector area and 0.9 for the total efficiency in the 150–1,500-keV range, the estimated source intensity is of the order of 5 photons ($cm^{-2}\,s^{-1}\,2\pi^{-1}$) in this energy range.

For a better understanding of the planet's characteristics, more information can be extracted from the spectra. A total spectrum including 116 individual spectra during the fourth pericentre passage is shown in Fig. 2 with 1σ errors bars. The measurement period started at 11:27:30 UT on 11 February 1989, that is 42 s before the pericentre time, and lasted 487 s. A cosmic spectrum normalized to the same duration, measured 1 h after the pericentre passage (between 12:27:09 and 12:36:50 UT) is plotted below the total spectrum for comparison and will be used as a background reference. The line near 4.5 MeV is due to the internal calibration source.

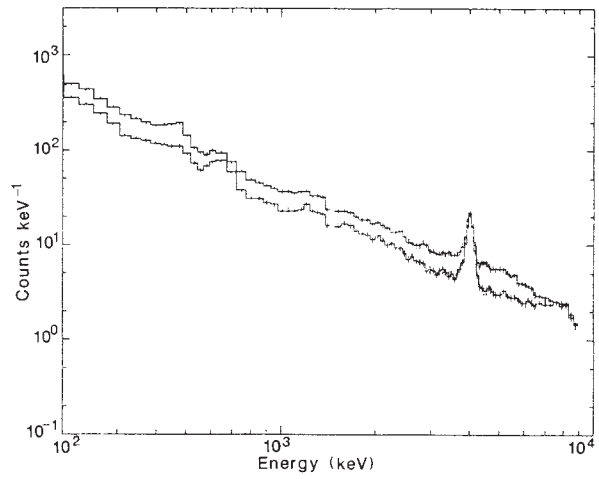


FIG. 2 Total spectrum registered at pericentre number 4. A spectrum measured 1 h later is plotted below for comparison. The line near 4.5 MeV in both spectra is caused by the internal calibration source.

This figure clearly shows that the emission from the planet covers an energy range from a few hundred keV to 8 MeV. There is no contribution above this upper limit as expected from predominantly γ-ray-induced emissions[4]. Another characteristic of the spectrum is that it is almost parallel to the background spectrum. Here again the increase factor is of the order of 1.8. Some features are observed above the continuum: a clear line at 511 keV corresponding to the annihilation process and some features due to the CsI crystal activation (mainly at 0.7 and 1.4 MeV in iodine). The fact that the activation peaks are not stronger in the pericentre spectrum seems to indicate that the neutron albedo does not contribute a large fraction to the observed increase of the count rate.

The particular contribution of the planet is obtained by subtracting the background spectrum from the pericentre spectrum (Fig. 3). It is clear that this additional signal consists mainly of a continuum. The only evident feature is the peak at 511 keV

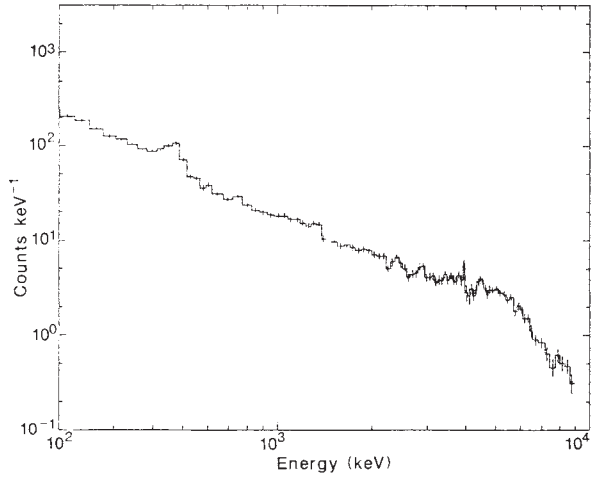


FIG. 3 Spectrum of the Mars planet contribution measured at pericentre 4. It shows a continuum above which an annihilation line contribution is seen at 511 keV and a broad contribution of probable nuclear lines between 4 and 7 MeV.

from the annihilation process. This line may be partly produced locally within the spacecraft body and the detector itself through albedo interactions, but it is also due to annihilation in the martian atmosphere and the surface. The relative amount of each possible line component will be established after a more careful study of the processes involved. Another important feature is the cutoff at 6 MeV with no significant contribution of the planet above ~8 MeV. However, as the count-rate spectrum in Fig. 3 tends to a constant level below 200 keV, the photon spectrum must be mainly above this value because the Compton effect is the dominant interaction in the detection process. This is similar to measurements of the lunar γ-ray albedo of which the energy spectrum is limited in the range from a few hundred keV to ~8 MeV (ref. 5).

When looking more carefully at Fig. 3, it seems that there is a 'shoulder' in the spectrum between ~4 MeV and ~7 MeV. This may indicate a possible mixed contribution of individual lines produced in the surface of Mars such as oxygen (lines at 4.438, 6.129, 6.917 and 7.117 MeV), iron (lines at 7.631 and 7.645 MeV), silicon (line at 4.934 MeV) or titanium (line at 6.762 MeV). The moderate energy resolution of the detector, the importance of the Compton effect in the detection process, as well as the role of the escape peaks and the bremsstrahlung process, makes line identification difficult. Further analysis needs an accurate procedure to unravel photon spectra from count-rate spectra. This task is under way. As this experiment allows a fine-time analysis, some temporal evolution already seen in the observed spectra may appear to originate in the longitudinal variations. If this hypothesis is confirmed, it will be possible to obtain information on the regional variability of the martian surface composition. □

Received 19 June; accepted 5 September 1989.

1. Surkov, Yu. A., Moskaleva, L. P., Manvelyan, O. S. & Khariukova, V. P. *Kosmich. Issled.* **XVIII**(4), 623-631 (1980).
2. Surkov, Yu. A. *et al. Nature* **341**, 595-598 (1989).
3. Sagdeev, R. Z. & Zakharov, A. *Nature* **341**, 581-585 (1989).
4. Reedy, R. C. in *Proc. 9th Lunar planet. Sci. Conf.* 2961-2984 (Pergamon, Oxford, 1978).
5. Bielefeld, M. J., Reedy, R. C., Metzger, A. E., Trombka, J. I. & Arnold, J. A. in *Proc. 7th Lunar Sci. Conf.* 2661-2676 (Pergamon, Oxford, 1976).

ACKNOWLEDGEMENTS. The authors wish to thank the Phobos Mission Operation Team for their part in the operation.

Vertical profiles of dust and ozone in the martian atmosphere deduced from solar occultation measurements

J. E. Blamont*, E. Chassefiere*, J. P. Goutail*, B. Mege*, M. Nunes-Pinharanda*, G. Souchon*, V. A. Krasnopolsky†, A. A. Krysko† & V. I. Moroz†

* Service d'Aeronomie du CNRS, BP No. 3, 91371 Verrieres-le-Buisson, France
† Space Research Institute, Profsoyuznaya 84/32, 117810 Moscow GSP-7, USSR

THE most sensitive parameter in a chemical model of the martian atmosphere is the vertical profile of ozone concentration, which has not previously been measured. Solar occultation measurements performed on the Phobos 2 spacecraft in martian orbit have provided us with such profiles, together with evidence for the existence of particles of small dimensions (0.1-μm diameter) in the 50- to 20-kilometre altitude range. These particles could be H_2O ice at the top of the observed range and dust at lower altitudes, possibly haematite, with a wide range of dimensions (down to nanocrystal size), dispersed in some other material (H_2O ice or silicates, for example). The relatively large number of these solid particles opens up the possibility of heterogeneous chemistry having a major role in the atmosphere of Mars.

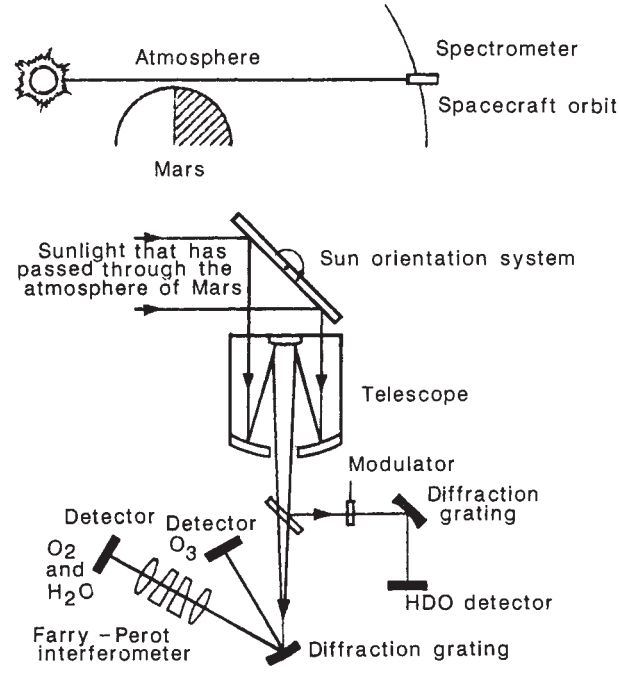


FIG. 1 Schematic diagram of instrument.

The instrument that we used to observe solar occultations in the atmosphere of Mars is shown in Fig. 1. A biaxial pointing device (a microprocessor-controlled flat mirror of elliptical shape, 6 × 8.5 cm, gimbal mounted, with coarse and fine adjustments forms the image of the Sun in the focal plane of a Cassegrain telescope (diameter 5 cm and equivalent focal length 75 cm). The pointing accuracy achieved in martian orbit was >20 arcsec, and the acquisition time was 4 s.

A diaphragm of 150-μm diameter placed in this focal plane reduces the angular dimension of the Sun to 1 arcmin and it is the spectrum of this spot that is analysed. The vertical resolution is 3 km in circular orbit but can fall to 10 km for observations from a high periapsis eccentric orbit or when the integration time constant is increased to 4 s.

The light admitted through the diaphragm is sent through a beam splitter into three different dispersive systems: a spectrograph using a holographic grating provides a spectrum of 512 elements in the spectral range 215-328 nm. It detects ozone in the Hartley band with a resolution varying from 1.2 nm at 215 nm to 2.4 nm at 328 nm; a Fabry-Perot interferometer placed behind a predispersing grating provides two spectra of 80 elements each, the first has a 0.2 nm spectral range centred at 760 nm (O_2 bands) and the second has a 0.3-nm spectral range centred at 936 nm (H_2O bands). The resolution is 150,000, the finesse is 50 and the spacing is 0.15 cm; a grating spectrograph provides two spectra of 24 elements each, in the ranges 5,290-5,370 cm^{-1} and 2,707-2,740 cm^{-1} for the detection of CO_2, H_2O and D_2O (see ref. 1). The resolution is 10^3.

The detectors of all spectrographs are linear arrays of photodiodes, which detect all the spectral elements simultaneously; the integration time varies from 0.1 to 4 s. The ultraviolet spectra that we report here were obtained using a nominal integration time of 0.1 s, with occasional use of a 0.4-s integration time.

The instrument operated between 8 February and 23 March. (the martian equinox occurred on 17 February; 32 occultations were observed during ingress and 1 was observed during egress. The lack of egress observations is because of the unstabilized altitude of the spacecraft during all but one eclipse. The first two occultation measurements were obtained when the space-

craft was in one eccentric orbit (periapsis, 4,270 km; apoapsis, 83,565 km; period, 79.2 h), the third when the spacecraft was in another eccentric orbit (periapsis, 9,800 km; apoapsis, 87,600 km; period, 86.5 h) and all other subsequent observations were made when the spacecraft was in a circular orbit (radius, 9,800 km; period, 8.1 h). (All distances given from the centre of Mars.) The latitude of the intersection of Sun-spacecraft axis with the surface of Mars varied from −11° on 8 February to +18° on 23 March, and the distance between the spacecraft and the limb was $\sim 10^4$ km.

Here we present only the results obtained with the ultraviolet spectrograph and the signal obtained in the near-infrared range integrated over the spectral range of 0.2 nm and 03 nm respectively for the 760 nm and 936 nm channels. The instrument had some defects. Stray light added an inverted solar spectrum to the data. (Extra intensity at 320 nm is due to stray light of wavelength 320 nm but the extra intensity at 220 nm is due to stray light with wavelengths in the range 500–400 nm.) Also there was a fast degradation of the optics in the range 260–230 nm after the shutter was opened on 5 February. The effect of stray light on the 220–320 nm range was removed by considering as the signal, the difference between the bottom of a Fraunhofer line and the adjacent maximum. The existence of stray light with a well-defined spectral composition has the favourable consequence of providing information on the occultations in the 320–520 nm range, however, with a poor resolution (10–50 nm depending on the position in the spectrum).

The most conspicuous feature in the data is the time lapse between the occultation curves in the visible (940 and 760 nm) and in the ultraviolet (320–260 nm): in the ultraviolet the extinction of solar light is observed at altitudes from 40 km down, whereas the extinction in the visible starts only around 30 km. We attribute this difference to the existence of solid particles whose dimensions d are on the order of the wavelength $\pi d/\lambda = 1$, that is, $d = 0.1$ μm, much smaller than the 4-μm particles

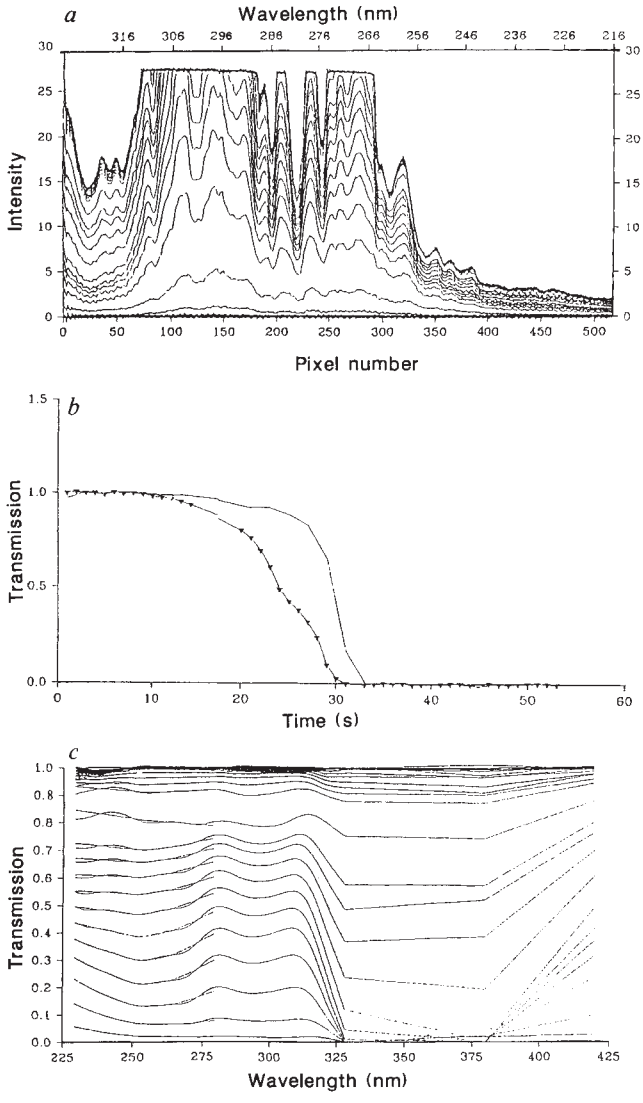


FIG. 2 Type I occultation curves obtained on 8 February 1989 (elliptic orbit). *a*, Series of solar spectra obtained during the occultation. Pixel 0 corresponds to $\lambda = 330$ nm, pixel 512 to $\lambda = 220$ nm. Note saturation at the beginning of the occultation. Pseudo absorption feature between pixels 0 and 70 is the result of shape of filter. *b*, Comparison of ultraviolet and visible occultation curves. Solid line is visible data, solid triangles are ultraviolet data. *c*, Extinctions in the range 225–280 nm, 280–310 nm and 310–330 nm are due to ozone absorption, solar-spectrum saturation artefacts and absorption band of unknown origin, respectively; the spectrum in the range 330–425 nm was deduced from stray-light analysis.

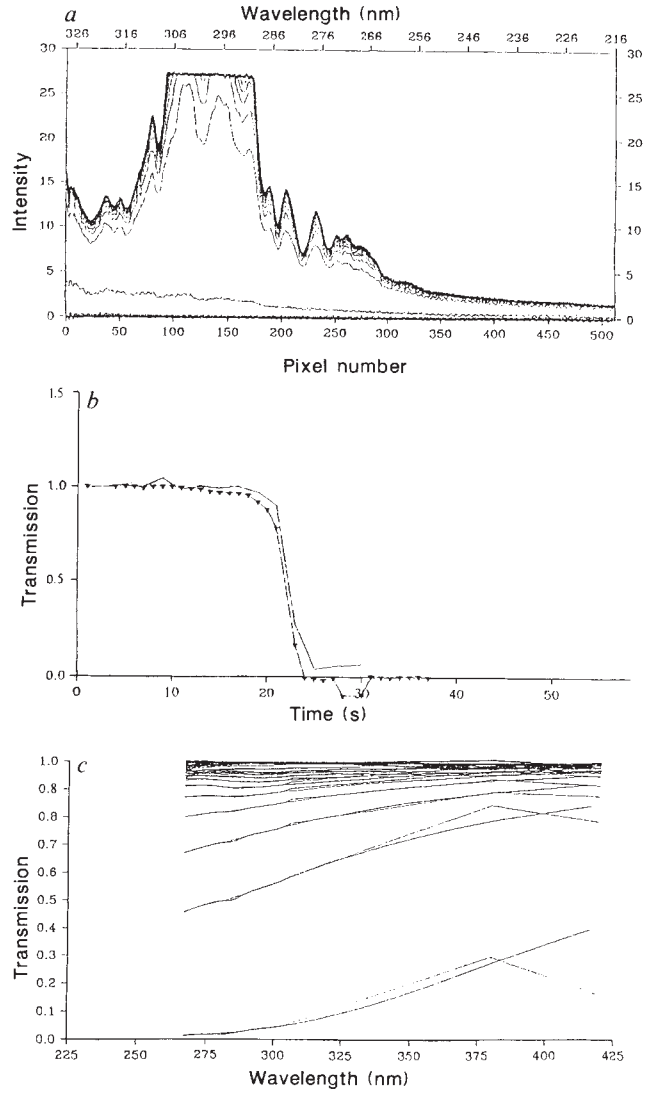


FIG. 3 Type II occultation curves (same presentation as Fig. 2 obtained on 16 March 1989 (circular orbit). *a*, Series of spectra obtained during occultation. Degradation of coatings by solar light results in loss of sensitivity. Compare lowest spectrum on Fig. 2*a* and 3*a* between pixels 0 and 150: there is evidence for absorption on Type I but not on Type II. *b*, Comparison of UV and visible occultation. Solid line is visible data; solid triangles are ultraviolet data. *c*, Series of extinction spectra. Solid lines are λ^{-4} theoretical curves. Dotted lines are experimental curves; note absorption above 400 nm.

detected by Viking at lower altitudes. We obtained no data below an altitude of 15–20 km. (The altitude determinations are preliminary and could be in error by ±5 km.)

In the ultraviolet range, the occultation curves are highly variable from orbit to orbit and we identify three types—24 curves of type I, 7 of type II, and 2 of type III. Type I (Fig. 2): the visible absorption starts 3–5 s (in circular orbit) after the ultraviolet absorption with a scale height of 5 km. The major characteristic of type I is the strong absorption in the spectum between 400 and 310 nm, usually below 40 km of altitude. In the majority of cases, the absorption between the altitudes of 50 and 40 km is greater at 280 nm than at 320 nm, but from 40 km downwards, it is greater at 320 than at 280. From 940 nm to 260 nm, the intensity of the absorbed light varies as $\lambda^{-1.5}$ indicating a mixure of Mie and Rayleigh scattering. Type II (Fig. 3): the extinction in the visible starts only 1 to 2 s after the ultraviolet. Both ultraviolet and visible transmissions decrease slowly for a short period, then very sharply: their scale height varies during the occultation from 7 to 2.5 km. There is no absorption in the 400–310 nm region. From 400 nm to 250 nm, the intensity of the light varies as $\lambda^{-3.8}$, indicating quasi-Rayleigh scattering. We suspect that the slope and the spectrum of these type II occultations may have been perturbed by a defect in the pointing system, and that the data are reliable only for the top of the atmosphere. Type III: the occultation curve is not monotonic, but shows a strong oscillation at an altitude of ~40 km. This indicates a thin cloud layer at this altitude, which removes the solar light at all wavelengths and perturbs the behaviour of the pointing system. Such horizontal layers have been frequently observed during the Viking mission.

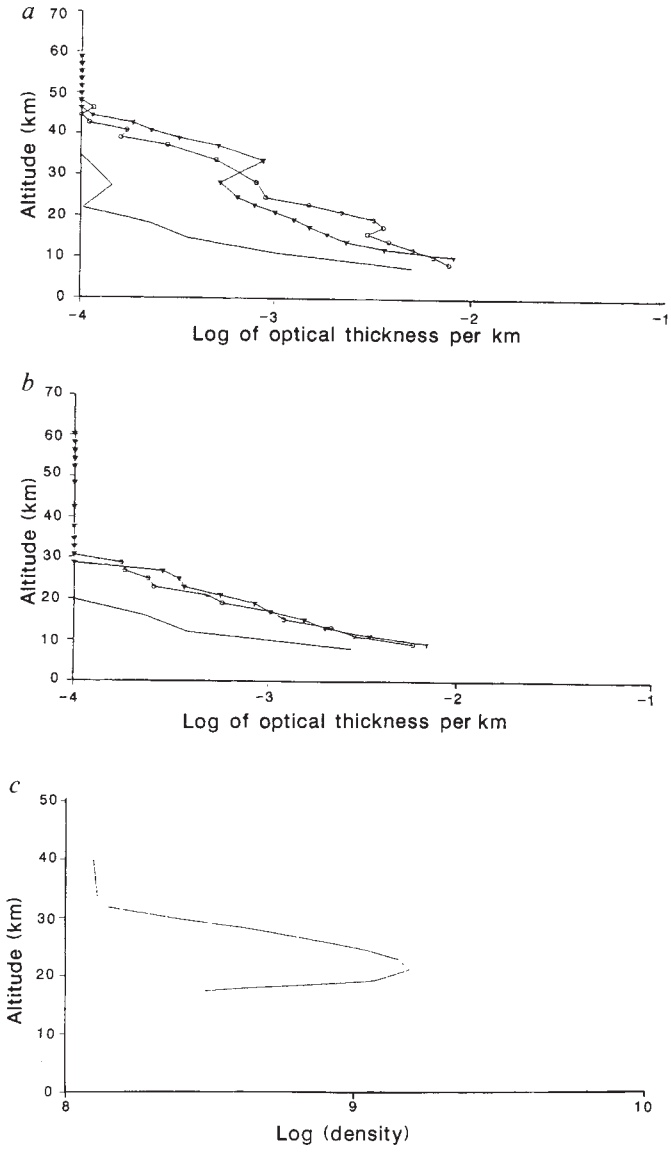


FIG. 4 *a*, Vertical distibution of extinction for type I on 8 February 1989. Open circles are 320-nm data; solid triangles are 280-nm data; solid line is the visible data. *b*, Another example of extinction for type I, 21 February 1989. *c*, Vertical distribution of ozone 8 February 1989.

Our observations are compatible with the existence of two kinds of particles, that were postulated based on the Mars 5 observations[2]:

Type A particles: Figure 2*c* gives the absorption spectrum characterizing these particles. From 230 to 330 nm, the spectrum is deduced directly from the observed spectrum and from 330 nm to 430 nm, we deduced the spectrum from an analysis of the stray light. In the latter case the resolution is poor but the intensities are reliable. The absorption increases sharply at 310 nm, peaks at ~380 nm and extends to ≥430 nm. The identification of the particles remains uncertain. Silicate dust originating from the soil is an obvious choice. Candidates for the metals in the dust, which are responsible for the ultraviolet absorption band, are Fe^{3+} (refs 3 and 4), TiO_2 (ref. 5), Mn^{2+} (ref. 6).

The particles detected in occultations seem to be different from the particles observed by the Viking lander. The spectra of those particles have a strong minimum in the imaginary part of the index of refraction at 760 nm with values of 0.025 at 760 nm and 0.050 at 936 nm[7], attributed to magnetite. The occultation curves presented here at these two wavelengths are nearly identical and present in nearly all the cases the same intensity for these two wavelengths. Particles of haematite, 0.1-μm dimension[8], might be responsible but the rapid decrease of absorption near 315 nm is not compatible with the presence of Fe^{3+}.

Type B particles: Figure 3*c* gives the extinction spectrum characterizing these particles. Because they possess no strong absorption features, they are not easy to identify. The atmospheric temperature is too high for CO_2 ice to exist, and therefore our prime candidate is H_2O ice. We have evidence for an absorption band starting at 380 nm which could coincide with the inflexion in the variation of the imaginary part of the refractive index of H_2O ice[9].

The optical thickness that results from the particles of either type is of the order of 10^{-4} km^{-1} at the altitude of 30–40 km and 10^{-3} km^{-1} at 20 km (Fig. 4*a*). This corresponds to an optical thickness of 10^{-8} cm^{-1} over a vertical column of 1 cm^2 section, or to a number density of the order of 10^4 particles cm^{-3} for an average radius of 5×10^{-2} μm. A molecule travelling with a thermal velocity will encounter such a particle every few seconds. Even with a small sticking coefficient, the molecules will be adsorbed within a short time, and reactions at the surface can occur that would be forbidden in an all-gaseous phase[10]. It is therefore necessary to modify the aeronomical models of CO_2 stability in the martian atmosphere, developed mainly in 1971 after the first Mariner 9 measurements of O_3 and CO[11,12]). *In situ* heterogeneous chemistry may have a major role in the martian atmosphere.

The occultations were performed in a dry atmosphere and in the equatorial region where even the total ozone density is below the detection threshold of the instruments previously used. Here we obtained ozone profiles during the whole period in which the instrument operated. From the absorption spectrum obtained on 12 February (Fig. 2*c*), we deduced the profile in Fig. 4*b*, assuming Rayleigh scattering by the atmospheric CO_2 (17% at 250 nm) and no other absorber in the spectral range. All the profiles obtained seem similar for type I occultations. Previously, only integrated values of the total column abundance of ozone had been obtained[13–15]. □

Received 12 June; accepted 10 August 1989.

1. Krasnopolsky, V. A. *et al. Nature* **341,** 603-604 (1989).
2. Krasnopolsky, V. A. & Parshev, V. A. *Planet. Space Sci.* **27,** 113-120 (1979).
3. Wagner, J. K., Hapke, B. W. & Wells, E. N. *Icarus* **69,** 14-28 (1987).
4. Egan, W. G. & Hilgeman, T. W. *The Optical Properties of Inhomogeneous Materials* (Academic, New York, 1979).
5. Pang, K. & Ajello, J. M. *Icarus* **30,** 63-74 (1977).
6. Hua, C. T., Dollfus, A. & Mandeville, J. C. in *Proc. 7th Lunar Sci. COnf.*, Vol. 3, 2605-2622 (Pergamon, New York, 1976).
7. Pollack, J. B. *et al. J. geophys. Res.* **84,** 2929-2945 (1979).
8. Morris, R. V. *et al. J. geophys. Res.* **94,** 2760-2778 (1989).
9. Grenfell, T. C. & Perovich, D. K. *J. geophys. Res.* **86,** 7447-7450 (1981).
10. Clark, I. D. *J. atmos. Sci.* **28,** 847-858 (1971).
11. Hunten, D. M. *Rev. Geophys. Space Phys.* **12,** 529-536 (1974).
12. Kong, T. Y. & McElroy, M. B. *Icarus* **32,** 168 (1977).
13. Barth, C. A. *et al. Science* **179,** 795-796 (1973).
14. Barth, C A. & Dick, M. L. *Icarus* **22,** 205-211 (1974).
15. Noxon, J. F., Traub, W. A., Carleton, N. P. & Connes, P. *Astrophys. J.* **207,** 1025-1035 (1976).

Solar occultation spectroscopic measurements of the martian atmosphere at 1.9 and 3.7 μm

V. A. Krasnopolsky, V. I. Moroz, A. A. Krysko, O. I. Korablev, V. S. Zhegulev, A. V. Grigoriev, A. Yu. Tkachuk, V. A. Parshev, J. E. Blamont* & J. P. Goutail*

Space Research Institute, Profsoyusnaya 84/32, 117810 Moscow, USSR
* Service d'Aeronomie, CNRS, 91370, Verrieres-le-Buisson, France

WE present solar occultation measurements, taken from Phobos 2, of the H_2O and CO_2 bands at 1.9 μm and of the dust absorption at 1.9 and 3.7 μm. From these, we have calculated vertical profiles of water vapour, dust and atmosphere temperature in the range of 10–70 km. The water-vapour mixing-ratio profile demonstrates oscillations from 10^{-5} to 10^{-3} with a period of 25 km. We explain this behaviour by the condensation and evaporation of water at minima and maxima of the temperature oscillations.

Our instrument was a grating spectrometer with a thermoelectrically cooled array of lead selenide detectors forming 48 pixels. The first-order diffraction covering the spectral range 2,740–2,707 cm^{-1}, was imaged onto 17 pixels and 20 pixels were used for the second-order diffraction, covering the spectral range 5,364–5,292 cm^{-1}. The orders of diffraction were separated by using appropriate filters and by not using 11 pixels in the middle of the array. The calibration data show that the resolving power of the instrument is close to the design specification of $\lambda/\Delta\lambda = 1{,}000$. We used the solar pointing system and telescope of the French solar occultation package[1]. Eight successful occultation observations were made at sunset, and one at sunrise, in the equatorial latitudes of Mars. Here we discuss the preliminary results of water vapour and dust measurements obtained from a sunset occultation observed on 13 March 1989. The interpretation of our HDO band (3.7 μm) observations will be given elsewhere.

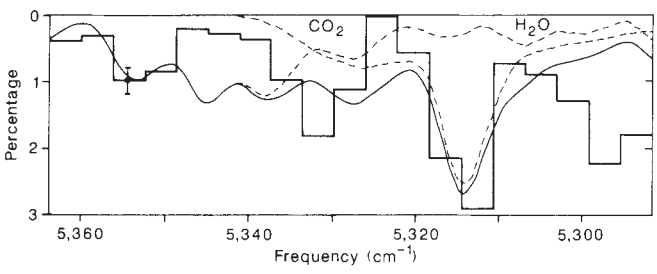


FIG. 1 Spectrum at 1.9 μm measured at 58 km (solid line) with error bars. Calculated absorption spectra of CO_2 and H_2O for the model atmosphere with $H_2O/CO_2 = 3\times10^{-4}$ are shown by dashed lines and their sum by a thin line.

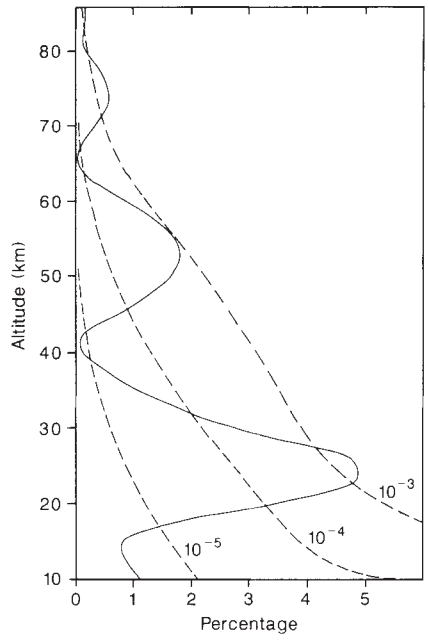


FIG. 2 Altitude profiles of the measured (solid line) and calculated (dashed line) water vapour absorption at 5,353 cm^{-1}. Calculations were for three H_2O mixing ratios, 10^{-5}, 10^{-4} and 10^{-3}.

One of the second-order spectra measured at an altitude of 58 km is shown in Fig. 1. Most of the features of this spectrum are typical of our spectra. Continuous absorption by aerosol is present in all spectra and constitutes ~20% in the spectrum in Fig. 1. The gaseous absorption readily identified is the CO_2 band at 5,314 cm^{-1}. The first four pixels on the left-hand side of the spectrum follow the spectrum of the H_2O band and we used these to evaluate the water-vapour mixing ratio. The calculated spectrum does not, however, reproduce the measured one well. The most important differences are the absence of the H_2O absorption at 5,340–5,350 cm^{-1} and the presence of features at 5,332 and 5,298 cm^{-1} in the observed spectra (these differences exceed the error bars, ±0.25%). The reason is probably the very low intensity of the gas absorptions (~1%) which is much less than the aerosol absorption. Small spectral variations of the latter may therefore significantly influence the measured spectrum.

We evaluated the intensity of the H_2O band at 5,353 cm^{-1} by comparing the signals for two pixels inside and two outside the band (Fig. 2). Three altitude profiles for this band that we calculated using three values of the water-vapour mixing ratio (10^{-5}, 10^{-4} and 10^{-3}) are also shown. At first glance the variation of the mixing ratio by two orders of magnitude is unexpected. Previous studies made by Viking spacecraft[2] showed that the water vapour is well mixed with the dust up to 10 km. This assumes a constant mixing ratio in this altitude range, which is below our range. Extrapolating the Viking data, one might suggest that the water-vapour mixing ratio on Mars should be

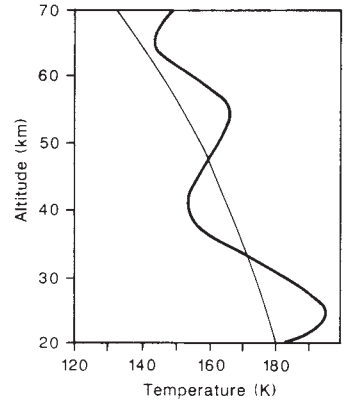


FIG. 3 Temperature profile of the mid-altitude atmosphere of Mars deduced from the data of Fig. 3. Thin line is the model profile[3].

constant with the mean value of $\sim 2\times10^{-4}$ up to 15–20 km and then decrease by one and two orders of magnitude by altitudes of 40 and 60 km respectively, following the saturated vapour-pressure curve. We suggest that the behaviour of the H_2O absorption is the result of the very strong temperature-dependence of the saturated water vapour pressure. The observed oscillations in the amount of water vapour correspond to the temperature oscillations with an amplitude of 10–15 K. Evaporation of ice aerosol grains occurs at temperature maxima and condensation of water vapour and formation of grains occurs at minima. To prevent sedimentation of the grains the atmospheric mixing must be strong, but to evaluate its lower limit the size distribution of ice grains must be known.

In Fig. 3 we compare the altitude profile of the atmospheric temperature deduced from the data in Fig. 2 with the model profile[3] given by the relationship $T(z)=185-0.05z-0.01z^2$. The length and the amplitude of the temperature oscillations

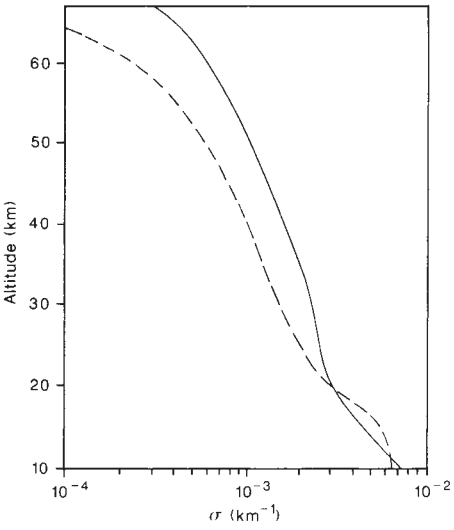


FIG. 4 Aerosol volume extinction coefficients at 1.9 and 3.7 μm (solid and dashed lines, respectively).

(25 km and 15 K) are close to those observed by the Viking entry probes[4]. These oscillations may be caused by internal-gravity waves (tidal waves). The sensitivity of water vapour pressure to the temperature of the atmosphere makes occultation measurements of water vapour bands a powerful tool in the study of these waves. But we do not see the increase of the amplitude of the wave with altitude that we expect for internal-gravity waves in the absence of dissipation.

Aerosol extinction can be measured easily by a solar occultation technique. The range of measurements covers from several per cent of absorption to ~99% absorption in the atmosphere. Figure 4 shows the Abel inversion of these measurements. At 60 km, the measured volume extinction coefficient at 1.9 μm is three times larger than that at 3.7 μm and becomes smaller than that at 3.7 μm between 20 and 10 km. Preliminary analysis of this behaviour in terms of Mie theory for two kinds of aerosol, water-ice particles and particles formed of a mixture of basalt and limonite (2.5%) (ref. 5), gives almost the same results: the effective mean radius of particles, r, is ~1 μm near 60 km, increasing to 3 μm at 10–20 km. A combination of these results with the ultraviolet and visible data[1] will allow further conclusions to be made about the aerosol composition and size distribution. Between 10 and 20 km the aerosol scale height measured at 1.9 μm is 11.5 km, (close to the atmospheric scale height). The equal scale heights presume well-mixed aerosol below 20 km. Extrapolating this mixing ratio to the surface of the planet we obtain an optical depth of aerosol equal to 0.3 which is in accord with the Mars 5 measurements[6] and Viking landers[7]. Above 20 km the aerosol mixing ratio seems larger than in the lower atmosphere. The problem should be considered, however, with the mean particle radius r taken into account. The sedimentation flux is roughly proportional to r^5, whereas the volume extinction coefficient varies as r^2. After this correction the ratio of the sedimentation flux to the atmospheric density is almost constant between 10 and 60 km. This may be important for theoretical study of the dust distribution in the martian atmosphere. □

Received 23 June; accepted 11 August 1989.

1. Blamont, J. E. *et al. Nature* **341**, 600–603 (1989).
2. Davies, D. W. *J. geophys. Res.* **84**, 2875–2879 (1979).
3. Krasnopolsky, V. A. *Photochemistry of the Atmosphere of Mars and Venus* (Springer, Berlin, 1986).
4. Seif, A. & Kirk, D. B. *J. geophys. Res.* **82**, 4364–4378 (1977).
5. Moroz, V. I. *Physics of Mars* (Nauka, Moscow, 1978; in Russian).
6. Krasnopolsky, V. A., Parshev, V. A., Krysko, A. A. & Rogachev, V. N. *Cosmic. Res.* **18**, 120–143 (1979).
7. Pollack, J. B. *et al. J. geophys. Res.* **82**, 4479–4496 (1977).

Magnetic fields near Mars: first results

W. Riedler*, D. Möhlmann†, V. N. Oraevsky‡, K. Schwingenschuh*, Ye. Yeroshenko‡, J. Rustenbach†, Oe. Aydogar*, G. Berghofer*, H. Lichtenegger*, M. Delva*, G. Schelch*, K. Pirsch*, G. Fremuth*, M. Steller*, H. Arnold*, T. Raditsch*, U. Auster†, K.-H. Fornacon†, H. J. Schenk†, H. Michaelis†, U. Motschmann †, T. Roatsch†, K. Sauer †, R. Schröter†, J. Kurths§, D. Lenners§, J. Linthe§, V. Kobzev‡, V. Styashkin‡, J. Achache‖, J. Slavin¶, J. G. Luhmann# & C. T. Russell#

* Institut für Weltraumforschung/ÖAW, 8010 Graz, Inffeldgasse 12, Austria
† Institut für Kosmosforschung, 1194 Berlin, Rudower Chaussee 5, GDR
‡ Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation, Troitsk, Moscow region, USSR
§ Heinrich-Hertz Institut, 1194 Berlin, GDR
‖ Institute de Physique du Globe, 4 Place Jussieu, 75252 Paris, Cedex 05, France
¶ Goddard Space Flight Center, Greenbelt, Maryland 20771, USA
\# Institute of Geophysics and Planetary Physics, University of California Los Angeles, Los Angeles, California 90024, USA

BASED on the examples of the Earth and Moon, we expect that Mars has or once had an intrinsic planetary magnetic field. The Earth has a large magnetic moment, 8×10^{15} T m^3, and the Moon's remanent magnetism in its surface rocks indicates that it once had a magnetic field[1]. The radius of Mars (3,390 km) is intermediate between that of the Earth (6,371 km) and of the Moon (1,738 km). Its smaller size has led to a more rapid thermal evolution than Earth's but we do not know whether Mars still generates a magnetic field or has passed into a Moon-like state. The Mariner 4, and Mars 2, 3 and 5 missions have placed upper limits on the martian magnetic moment[2,3] in the range of $1–5\times10^{12}$ T m^3 corresponding to an equatorial surface field of 25–125 nT. We have measured magnetic fields with high temporal resolution in the tail and down to an 850-km altitude. During four successive highly elliptical orbits we identified the position of the bow shock as well as of a transition layer, the 'planetopause'. Subsequent circular orbits at 6,000-km altitude provided the first high-resolution data in the planetary tail and indicate that the interplanetary magnetic field mainly controls the magnetic tail. We also detected magnetic turbulence when the spacecraft crossed the orbit of the moon Phobos, indicating the possible existence of a torus near the orbit of this moon.

Phobos 2 carried two triaxial flux-gate magnetometers: the experiment MAGMA (magnetic fields near Mars) and the

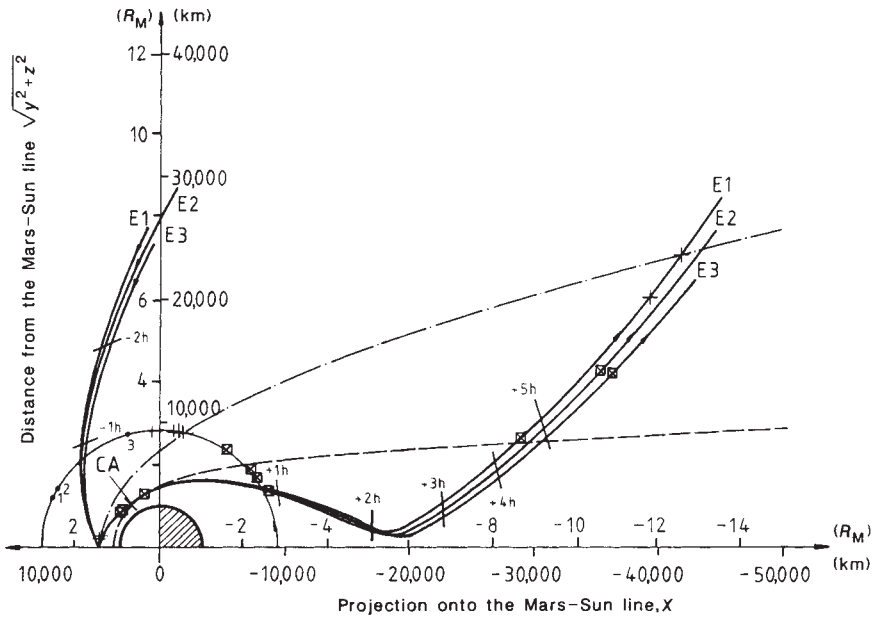


FIG. 1 Main boundaries and events observed by the FGMM and MAGMA experiments during the first three low-altitude elliptical orbits (E1, E2 and E3) and during circular orbits (outer semi-circle) around Mars. Tick marks indicate hours relative to the closest approach (CA). Squares denote inbound and outbound crossings of the planetopause, crosses the positions where the magnetometers observed the bow shock.

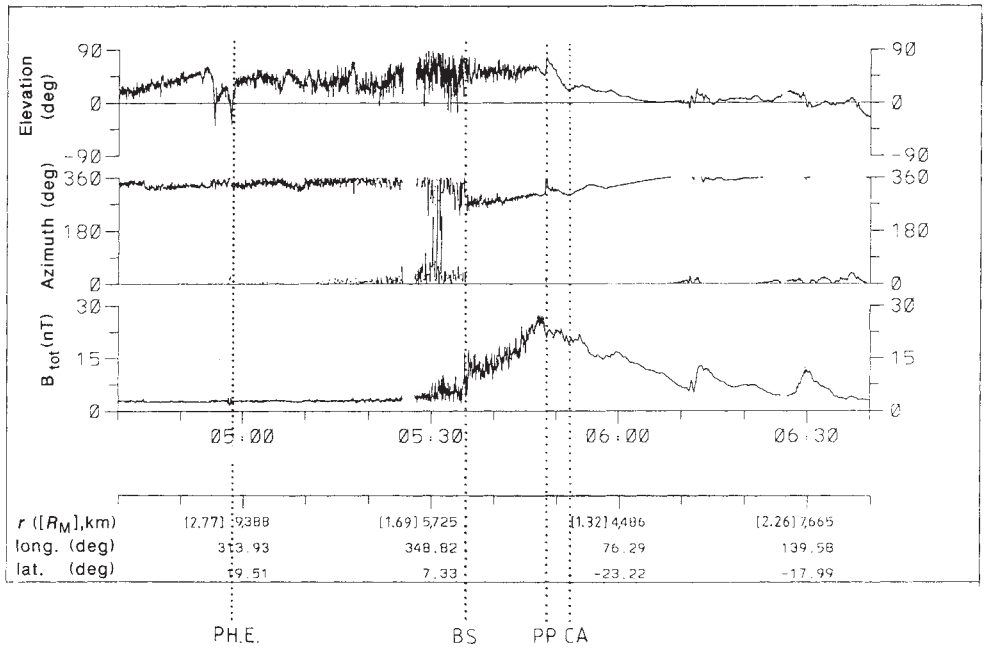


FIG. 2 Magnitude and orientation of the magnetic field vectors (FGMM experiment) measured during elliptical orbit 3 near closest approach (altitude = 850 km, $r = 1.25\ R_M$). Temporal resolution is 2.5 s. The magnetic field is presented in the areocentric solar-ecliptic coordinate system, where the elevation is measured with respect to the orbital plane of Mars (x, y) and the azimuth with respect to the Mars-Sun line (azimuth is 0 towards the sun). The lower panel shows longitude and latitude and the distance of the spacecraft from the centre of Mars r. BS denotes the inbound bow shock, PP the inbound planetopause and CA the closest approach. The disturbance observed at 4:48 UT (label PH.E.) is the so-called Phobos event.

experiment FGMM (flux-gate magnetometer Mars). Both instruments had a range of ±100 nT, a resolution of 0.05 nT and returned data at the rate of one vector every 1.5, 2.5, 45 or 600 s depending on the telemetry mode of the spacecraft. The MAGMA magnetometer was mounted on the top of a 3.5-m boom and the FGMM magnetometer was mounted on the same boom, 1 m closer in. The dual-magnetometer configuration and the pre-launch magnetic mapping of the spacecraft allowed us to estimate the spacecraft field at the location of the MAGMA sensor (<1 nT). The total offset (sum of the internal offset and the spacecraft field) was calculated from the magnetic field data from the spinning and rolling sessions using statistical methods[4]. More details about the instruments and the in-flight performance can be found elsewhere[5].

Figure 1 shows the first three elliptical orbits of Phobos 2 and the general features of the solar-wind interaction with Mars in cylindrical coordinates. Initially, Phobos 2 entered a highly elliptical orbit with a 77-h period and a closest approach of $r = 1.2507\ R_M$ (measured in units of Mars radii from the centre of the planet). The martian tail was crossed 2.4 h after closest approach at a radial distance of $\sim 5.6\ R_M$. In Fig. 1 the distance of the spacecraft from the Mars-Sun line is plotted against the distance along it; the orbit is therefore folded over on itself. After four low-altitude orbits of this type the spacecraft performed one elliptical orbit with a periapsis at 2.83 R_M and later transferred into a circular orbit at 2.83 R_M, 300 km above the orbit of Phobos, circling the planet every 8 h. Until the end of the mission on 27 March 1989 we obtained magnetic field data during 80 circular orbits—50 orbits with a resolution of 45 s and 30 orbits with a resolution of 600 s. Near the pericentre of the four low-altitude elliptical orbits, the magnetometers transmitted 1 vector every 1.5 s (MAGMA) and 2.5 s (FGMM). Figure 2 shows magnetic field measurements obtained on the third elliptical orbit near the pericentre. We observed a typical undisturbed interplanetary magnetic field (IMF) with a magnitude B_{tot} of ~ 3 nT and with fluctuations of the field direction (azimuth, elevation) until 5:30 UT when the field magnitude started to become turbulent. Five minutes later when Phobos 2 crossed the bow shock we observed a jump in the field magnitude (to a maximum of 28 nT) and a decrease in the fluctuations of the

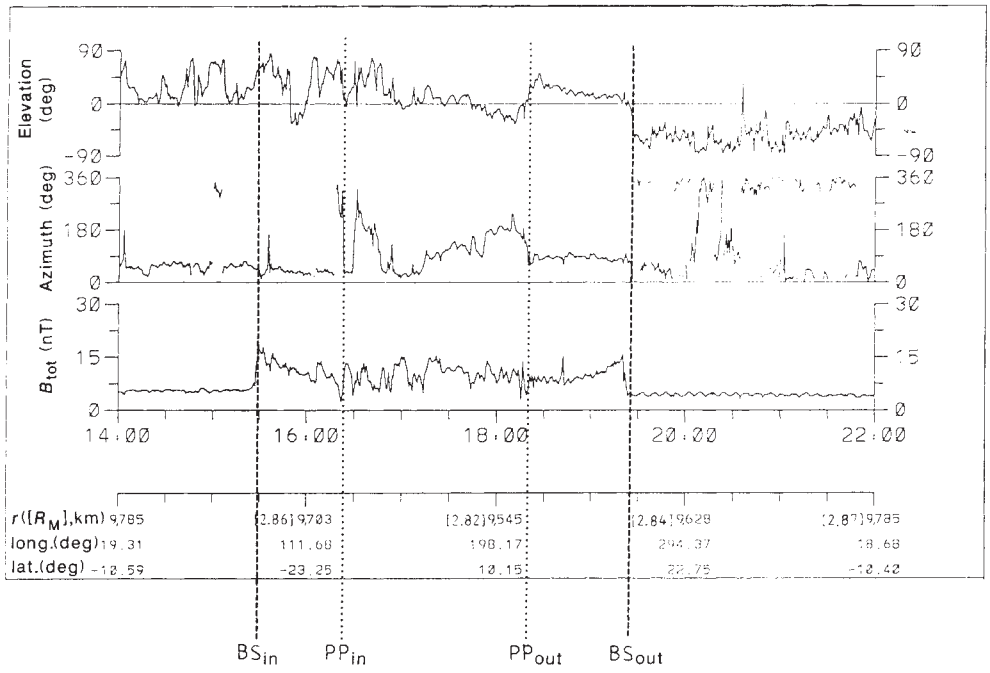


FIG. 3 Magnitude and orientation of the magnetic field (MAGMA experiment) during one of the circular orbits. Temporal resolution is 45 s. BS_{in}, PP_{in}, PP_{out} and BS_{out} are inbound and outbound bow shock and planetopause, respectively.

azimuth and elevation. Then the spacecraft passed a boundary, which we call the planetopause, that is characterized by a rotation and by a drop in turbulence of the magnetic field. At the planetopause the solar-wind protons also disappeared[6]. After this boundary, the field remained quiet and the magnitude decreased gradually. We call the disturbance at 4:48 UT near the orbit of the moon Phobos, the Phobos event. We observed the same boundaries and events during other elliptical orbits; their time and position (areocentric distance, longitude and latitude) are summarized in Table 1. Figure 3 shows the magnetic fields we observed during one of the circular orbits. The bow shock, inbound and outbound is characterized by a jump of the total magnetic field. The reversal of B_x at 17:20 UT is typical for the crossing of the martian tail and the polarity of the reversal depends mainly on the orientation of the perpendicular, undisturbed IMF component observed before the bow shock. The inbound and outbound crossings of the planetopause are characterized by the disappearance and reappearance respectively, of the solar-wind protons[6]. In some cases the magnetic field magnitude increased or decreased during the inbound or outbound crossings, respectively.

As illustrated in Fig. 1 the orbits of Phobos 2 were better suited for investigating the solar-wind interaction with Mars than those of previous spacecraft[7]. The elliptical orbits came almost twice as close as the orbits of previous spacecraft and provided data directly behind the planet. The location of the bow shock has been studied extensively on previous missions. The analysis[7] had obtained subsolar (nose) radii of 1.36–1.85 R_M and terminator radii of 2.36–3.07 R_M. Initial Phobos 2 observations show that the bow shock is located with a nose radius of ~1.45 R_M and a terminator radius of about 2.40 R_M We calculated these values using a gas-dynamic model[8] with the shape of the obstacle from Fig. 1 (dashed line), a Mach number of 10 and a ratio of specific heats of 5/3. We fitted the obstacle to

TABLE 1 Timing and position of boundaries and events

Orbit	Inbound Bow shock	Inbound Planetopause	Outbound Planetopause	Outbound Bow shock	Phobos event	Closest approach
E1						
1, 2 February 1989	18:25	18:37	23:15	3:00–4:10	17:45	18:39:37
r (km)	4,984	4,278	30,264	44,204–47,774	9,711	4,260
long., lat. (deg)	7, 1	42, −14	197, 4	206, 8–208, 9	315, 20	51, −17
E2						
4, 5 February 1989	00;01	00:22	06:50	no data	23:20	00:15:47
r (km)	4,997	4,400	38,175		9,855	4,252
long., lat. (deg)	5, 1	71, −22	201, 6		313, 20	49, −17
E3						
8 February 1989	05:35	05:49	12:40	no data	04:58	05:51:58
r (km)	5,199	4,277	39,015		9,634	4,244
long., lat. (deg)	359, 3	37, −14	200, 7		312, 20	48, −17
K1						
2 March 1989	15:27	16:21	18:20	19:23		
r (km)	9,749	9,669	9,539	9,577		
long., lat. (deg)	85, −25	128, −20	213, 16	264, 25		
K2						
11 March 1989	00:13	01:06	02:36	03:50		
r (km)	9,687	9,700	9,694	9,670		
long., lat. (deg)	101, −24	141, −13	203, 14	261, 26		

Times are in UT and distances are measured from the centre of Mars. E1, E2 and E3 denote the first three of five elliptical orbits; K1 and K2 are examples of the about 80 circular orbits from which magnetic field data are available. The boundaries and events are determined with an accuracy of ~1 min.

the observed locations of the planetopause, where the solar-wind protons are disappearing. The term planetopause was chosen by analogy with the cometary cometopause where the solar-wind ions also disappear[9]. The nose of the fitted planetopause is 0.1180 R_M (400 km) above the martian surface. As the calculated bow shock (dash-dotted line in Fig. 1) fits the observed bow-shock crossings well (crosses in Fig. 1), we conclude that the observed planetopause is close to the obstacle that deflects the solar-wind flow around Mars thus forming the bow shock.

The wake of a magnetized planet is a region of magnetic fields with a large sunward or antisunward component. A current sheet separates the regions of opposite magnetic polarity or tail lobes. The polarity is controlled only by the intrinsic magnetic moment of the planet. By contrast an induced magnetotail has lobes with polarities controlled by the interplanetary magnetic field. We therefore checked the influence of the interplanetary field on the polarity of the lobes during the circular orbits and found several cases similar to Venus[10] where a reversal of the perpendicular IMF component caused a reversal of the tail lobes. This indicates the existence of an ionospheric type of interaction with the solar wind at distances $>2.8\ R_M$.

The analysis of the magnetic field data that we have done so far gives no conclusive evidence for an intrinsic magnetic field. A future analysis of the Phobos 2 magnetic field and plasma data near the closest approach ($r = 1.25\ R_M$) together with computer simulations of the interaction of Mars with the solar wind will improve our knowledge about the parameters of the intrinsic field.

Although we have no evidence for any direct effect of the moon Phobos on the solar wind we have observed an occasional disturbance in the magnetic field as the spacecraft crossed the moon's orbit (labels 1, 2 and 3 in Fig. 1 denote the position of the moon). A possible explanation is that the solar wind interacts with a torus of dust generated by meteorite bombardment of the surface of Phobos[11]. The disturbance is, however, not unlike disturbances caused by particles backstreaming from planetary bow shocks so these observations need further investigation. □

Received 19 June; accepted 15 September 1989.

1. Runcorn, S. K. *Scient. Am.* December, 60–68 (1987).
2. Dolginov, Sh. Sh. *Earth Moon Planet.* **37,** 17–52 (1987).
3. Russell, C. T. in *Geomagnetism* (ed. Jacobs, J. A.) Vol. 2, 457–523 (Academic, New York, 1987).
4. Hedgecock, P. C. *Space Sci. Instrum.* 83–90 (1975).
5. Aydogar. Oe. *Austrian Acad. Sci., Report IWF*-8904 (1989).
6. Rosenbauer, H. *et al. Nature* **341,** 612–614 (1989).
7. Vaisberg, O. L., Luhmann, J. G. & Russell, C. T. *J. geophys. Res.* (in the press).
8. Spreiter, J. R., Summers, A. L. & Rizzi, A. W. *Planet. Space Sci.* **18,** 1281–1299 (1970).
9. Gringauz, K. I. *et al. Geophys. Res. Lett.* **7,** 613–616 (1986).
10. Yeroshenko, Ye. *Cosmic Res.* **17,** 77–87 (1979).
11. Ip, W. *Icarus* **76,** 135–145 (1988).

First measurements of plasma waves near Mars

R. Grard*, A. Pedersen*, S. Klimov†, S. Savin†, A. Skalsky†, J. G. Trotignon‡ & C. Kennel§

* Space Science Department, European Space Agency/ESTEC, Postbus 299, 2200 AG Noordwijk, The Netherlands
† Space Research Institute, Profsoyuznaya 84/32, 117810 Moscow GSP-7, USSR
‡ Laboratoire de Physique et Chimie de l'Environnement CNRS, 45045 Orléans, France
§ Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024, USA

HERE we report preliminary results from electic field measurements in the environment of Mars using the plasma-wave system on board the Soviet spacecraft Phobos 2. It also includes a Langmuir probe which measured plasma densities. Electron-plasma oscillations observed upstream of the bow shock correspond to a solar-wind density of 2 cm^{-3}. The shock-foot boundary was crossed up to three times on each orbit. The shock ramp was detected at altitudes between 0.45 and 0.75 Mars radii (R_M) above the planetary surface. The density increased by about a factor of two at the ramp. The shock position, although variable, seemed to be consistent with previous measurements. The downstream magnetosheath contained broadband electric-field noise below the plasma frequency. The boundary of the obstacle, or planetopause, was crossed at altitudes of the order of 0.28 R_M; the cold plasma density was highly variable within the planetopause and reached the unexpected value of 700 cm^{-3} on the third orbit, at 0.25 R_M altitude. Bursts of waves with frequencies below the electron cyclotron frequency, possibly in the whistler mode, occur within the planetopause.

The scientific objectives of the plasma wave system (PWS) on board Phobos 2 are: (1) to analyse the frequency spectrum of the electromagnetic and electrostatic waves in the environment of Mars; (2) to detect the boundaries associated with the interaction of the solar wind with the ionosphere and magnetosphere of the planet; (3) to measure the plasma density distribution in the martian ionosphere; (4) to search for the possible existence of plasma density enhancement and wave activity resulting from the interaction between the solar wind and any material possibly outgassed from the Martian satellite Phobos; (5) to measure the spacecraft potential. The PWS instrument includes two sensors: a dipole antenna made of two 10-cm diameter spheres separated by 1.45 m and a Langmuir probe with a collecting area of 6 cm^2. Further technical information about the instrument is given elsewhere[1]. Electrostatic plasma waves were detected sporadically in the upstream solar wind on all four passes within an areocentric distance of 4.8 R_M. Because the frequency responses of adjacent filters in the PWS crossover at attenuations of only 3 dB, an intense narrow-band signal is detectable in several frequency channels. By assuming that the plasma frequency corresponded to the highest intensity channel, however, we deduced electron-plasma frequencies and densities consistent with those expected. The observations made during the first approach to the planet (Fig. 1, panel 1) yield a plasma frequency and electron density of 13 kHz and 2 cm^{-3}, respectively.

Phobos 2 crossed the outer boundary of the shock foot at 18:08, 18:12, and 18:14 UT on the first orbit (Fig. 1, panel 2). This boundary is the envelope of turnaround distances of solar-wind ions that are specularly reflected at, and are subsequently returned to, the shock ramp after completing a partial gyration in the upstream magnetic field[2]. The foot boundary was identified by a change in character of both the wave and Langmuir-probe data. As at Earth and Jupiter[3], the plasma waves in the shock foot of Mars appear as broad-band noise extending from below the lower hybrid frequency to the electron-plasma frequency. The Langmuir probe detected a slight gradual increase in electron density throughout the shock foot.

The shock ramp, which separates the shock foot upstream from the magnetosheath downstream, can be more easily identified from changes in overall Langmuir-probe response than they can from the broadband plasma wave shown in Fig. 1. More detailed inspection of the plasma-wave spectra indicates that the electron density jumped by about a factor of two at the shock ramp. A more precise evaluation of this density jump will require further analysis. The shock foot and magnetosheath are both regions of enhanced plasma wave activity with similar spectra, except that electron-plasma oscillations were more apparent upstream.

We have chosen the neutral name 'planetopause' (instead of magnetopause or ionopause) for the boundary separating the shocked magnetosheath from the planetary obstacle. Planetopause crossings were identifiable in both wave and Langmuir-probe data. Inbound crossings were preceded by significant modifications in the wave spectra, illustrated by the fourth panel

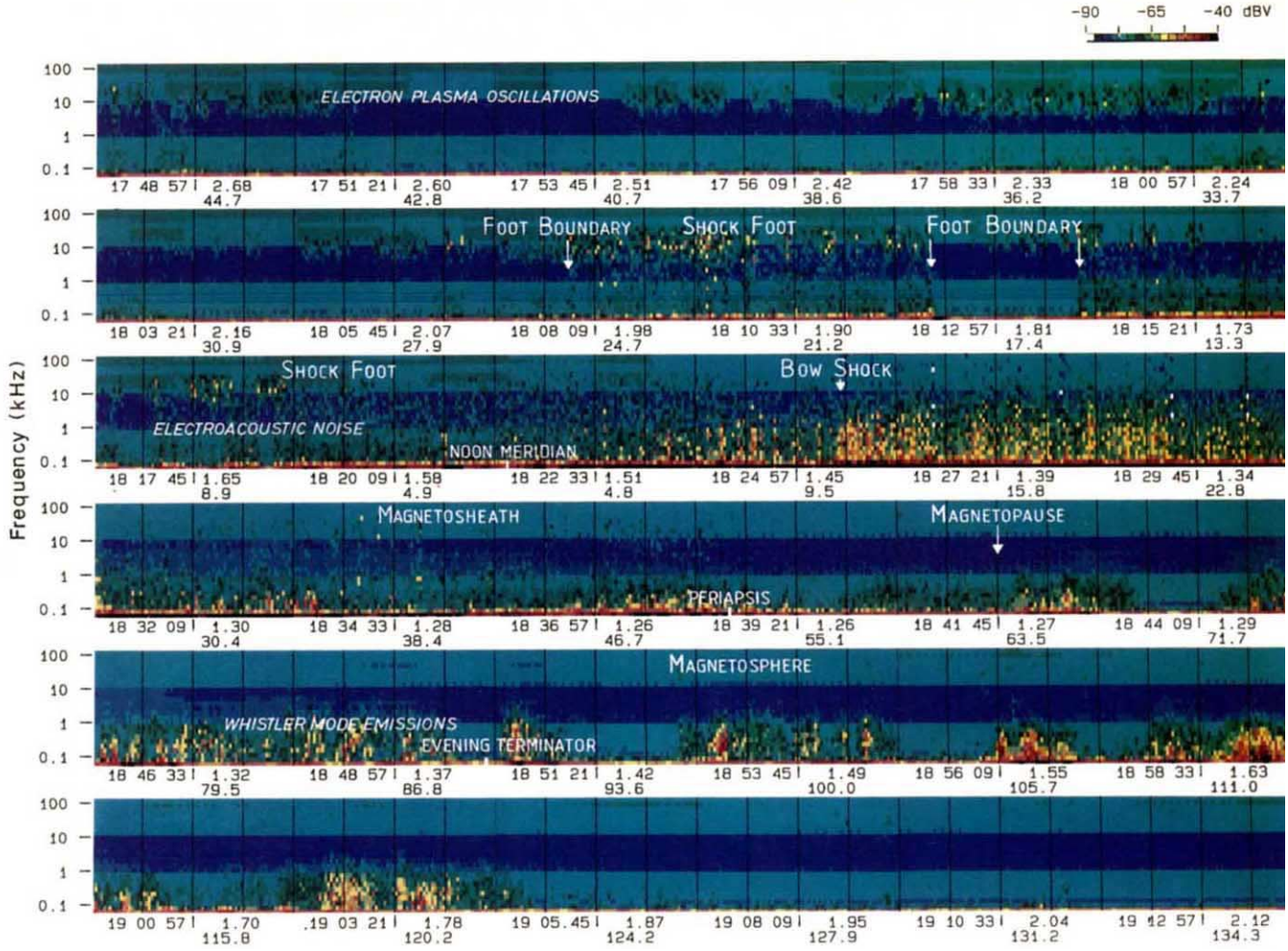


FIG. 1 Summary of a continuous-wave data sequence recorded on 1 February 1989 with the dipole antenna of PWS during the first orbit of Mars by Phobos 2. This spectrogram is a raster made of the 25 signals delivered by the filter bank; the r.m.s. value of the voltage measured with each filter is evaluated with the colour scale at the top right (−40 dBV = 10 mV r.m.s.). The central frequency of the bandwidth covered by each channel is marked on the left axis. Time (UT) is given along the horizontal axis to the left of each tick; the areocentric distance of the spacecraft (R_M) and the solar zenith angle (deg) are written to the right of each tick, top and bottom respectively.

of Fig. 1. The broad-band, presumably electrostatic, noise disappeared before Phobos 2 left the magnetosheath. The Langmuir-probe data changed character after every planetopause crossing; the temporal profiles were smoother, and the densities were higher inside the planetopause. The density just inside the planetopause varied from orbit to orbit.

The locations of the shock-foot boundary (open circles), shock ramp (crosses) and planetopause (triangles) observed in the sunward hemisphere during each of the four orbits are shown in Fig. 2. With one exception, the ramp crossings are consistent with a model summarizing earlier measurements of the bow shock[4] of Mars; we do not know whether the distinction between foot and ramp was drawn in previous data. One ramp crossing occurred somewhat further from the planet than all others detected thus far, and so did its shock-foot boundary. The thickness of the shock foot and magnetosheath were comparable, several tenths of R_M, confirming the prediction[5] that the foot is a prominent feature of the interactions of the martian environment with the solar wind.

The plasma density was different at each planetopause crossing and was highly variable inside the planetopause. Phobos 2 encountered regions of enhanced density, typically 100 km in extent along the orbit, in which plasma wave activity was not observed. Waves occurred primarily in discrete bursts of 0.5–2 min duration and 100–400 km extent. The signals in the burst between 19:02:30 and 19:04:00 UT (Fig. 1, panel 6) had frequencies <1 kHz, the electron cyclotron frequency in a 36-nT magnetic field. This suggests, but in the absence of wave magnetic-field data does not prove, that the bursts are in the whistler mode.

Phobos 2 measured the highest plasma density, 700 cm^{-3}, near periapsis on the third orbit (Fig. 3); at that time the electron temperature was ~1 eV. Assuming thermal equilibrium, the inferred plasma pressure was that developed by a 25-nT magnetic field, comparable with typical values of the solar-wind dynamical pressure; this peak density occurred 7 min after planetopause crossing and was about twice that observed at this boundary. The peak density on the third orbit was a factor of two or more larger than those encountered on the other orbits, indicating that the plasma density in the dayside environment of Mars varies from day to day. The third-orbit periapsis density is unusually large but is not the result of any instrumental effect or spacecraft interference. The retarding-potential analysers on board the Viking landers[6] measured comparable densities, but at much lower altitudes (300–350 km) than those of the Phobos 2 orbit, whose periapsis altitude was 860 km.

Correlative studies that combine magnetic field, plasma, and energetic-particle data are needed to understand the role played by thermal plasma in the Mars/solar-wind interaction, to determine whether the planetopause is an ionopause, a magnetopause

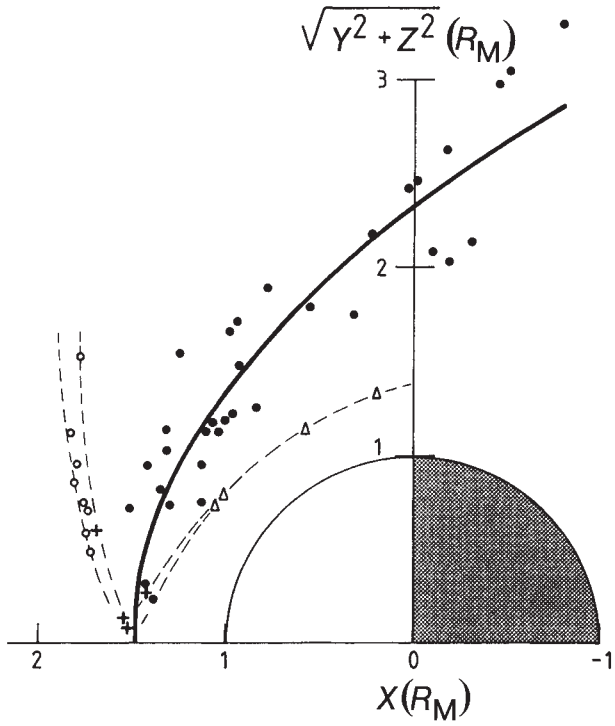


FIG. 2 Locations of the martian shock-foot boundary, shock ramp and planetopause observed with Phobos 2 (open circles, crosses and triangles, respectively, between the envelopes, dashed lines, of the four orbits). The average location of (presumably) the shock ramp (thick line) based on Mars 2, 3 and 5 and Viking 1 observations (full circles) is shown for comparison in aberrated coordinates[4].

or a mixture, and to estimate the convection rate inside the martian magnetosphere[7].

Plasma density and wave-activity enhancements have also been observed in the solar wind when Phobos 2 was at an areocentric distance of 2.77 R_M, which corresponds precisely to the orbital radius of one of the two natural satellites of Mars,

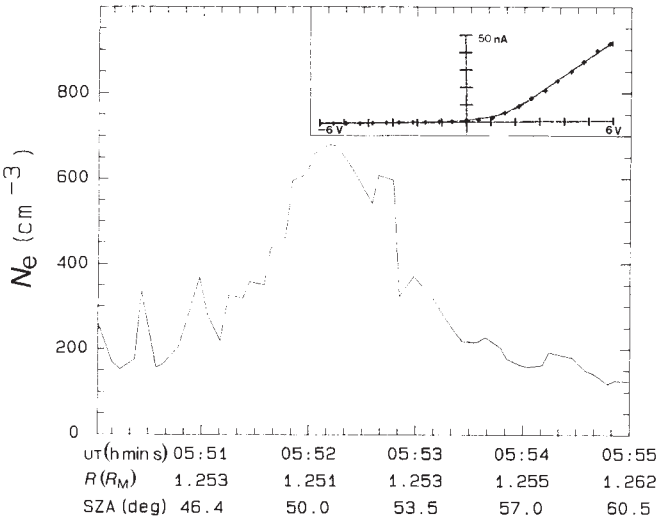


FIG. 3 Electron density profile recorded on 8 February 1989 when Phobos 2 was crossing the periapsis of the third orbit around Mars. Universal time (UT), areocentric distance and solar zenith angle (SZA) are given along the horizontal axis. The inset in the upper-right corner shows a typical current response of the Langmuir probe recorded at 05:52 UT when the bias voltage is swept between −6 V and +6 V with respect to the spacecraft structure; experimental points are represented by crosses; the continuous curve is a theoretical fit to the measurements from which the plasma parameters are derived.

its namesake. Further findings may be expected, because these data represent only a small fraction of the information that was transmitted by Phobos 2 during its two-month operation in the martian environment. □

Received 19 June; accepted 9 August 1989.

1. Grard, R., Klinge, D., Klimov, S., Savin, S. & Trotignon, J. G. *J. Phys. E* **22**, 656–662 (1989).
2. Gosling, J. T. & Thomsen, M. F. *J. geophys. Res.* **90**, 9893–9896 (1985).
3. Moses, S. L. *et al. Geophys. Res. Lett.* **12**, 183–186 (1985).
4. Slavin, J. A., Holzer, R. E., Spreiter, J. R., Stahara, S. S. & Chaussée, D. S. *J. geophys. Res.* **88**, 19–35 (1983).
5. Moses, S. L., Coroniti, F. V. & Scarf, F. L., *Geophys. Res. Lett.* **15**, 429–432 (1988).
6. Hanson, W. B., Sanatani, S. & Zuccaro, D. R. *J. geophys. Res.* **82**, 4351–4763 (1977).
7. Kennel, C. F., Coroniti, F. V., Moses, S. L. & Zelenyi, L. M. *Geophys. Res. Lett.* **16**, 915–918 (1989).

ACKNOWLEDGEMENTS. The PWS team would like to express their appreciation for the contribution made by Fred Scarf, who died in Moscow, a few days after the launch of Phobos. We thank D. Klinge, A. Butler, K. Hjortnaes, H. Christensen, H. Berle, J. Bouman, A. Fransen, J. Heida, R. Scheper, H. Hauser, J. M. Moreau, Z. Krawczyk, K. Dudzinski, M. Sosnowsky, G. Avenesova, S. Romanov and N. Rybyeva. C.K. acknowledges the financial support of NASA.

First measurements of the ionospheric plasma escape from Mars

R. Lundin*, A. Zakharov†, R. Pellinen‡, H. Borg*, B. Hultqvist*, N. Pissarenko†, E. M. Dubinin†, S. W. Barabash†, I. Liede‡ & H. Koskinen‡

* Swedish Institute of Space Physics, PO Box 812, S-98128 Kiruna, Sweden
† Space Research Institute, Profsoyuznaya 84/32, 117810 Moscow GSP-7, USSR
‡ Finnish Meteorological Institute, PO Box 503, SF-00101, Helsinki, Finland

BEFORE the Phobos mission, the magnetosphere of Mars was relatively unexplored. Basically all data on the charged-particle environment of Mars had come from the Soviet Mars missions (Mars 2, 3, 4 and 5). Here we report the results of an ion-composition experiment on board the Soviet Phobos 2 spacecraft which allowed us to determine the loss of plasma from the ionosphere of Mars. Because Mars has a very small intrinsic magnetic field, the ion outflow had been expected to be similar to that of Venus. Surprisingly, there were many similarities between the ionospheric outflow from Mars and the Earth. The ion loss from Mars results from both ion pick-up that results from mass-loading of the solar wind in the Martian boundary layer and ionospheric O^+ beams of energies up to several keV, possibly from upward acceleration processes similar to those observed above the Earth's auroral oval. A preliminary estimate of the ionospheric outflow from Mars indicates that the planet is losing oxygen at a rate of $\sim 3\times10^{25}$ ions s^{-1}. This corresponds to an evacuation of its present total atmospheric oxygen content (contained in CO_2 and O_2) in less than 100 million years.

Previous missions to Mars have shown (see, for example, ref. 1) that the interaction with the solar wind forms both a bow shock and a magnetopause. There is[2] a smooth transition from the shocked solar-wind plasma to the hot plasma cushion above the ionosphere, which closely resembles the solar-wind interaction with Venus, a planet without an intrinsic magnetic field. A boundary layer inside the magnetosheath with ion fluxes of lower energies than in the magnetosheath has also been observed[2], possibly indicating an ionospheric origin of these ions.

The contribution of the solar-wind interaction to the dehydration of Mars has been discussed[2,3]. The lack of a strong intrinsic magnetic field on Mars (and on Venus) may have had a major role in the loss of atmosphere (and hydrosphere) on Mars by the direct exposure of the upper ionosphere to the solar wind.

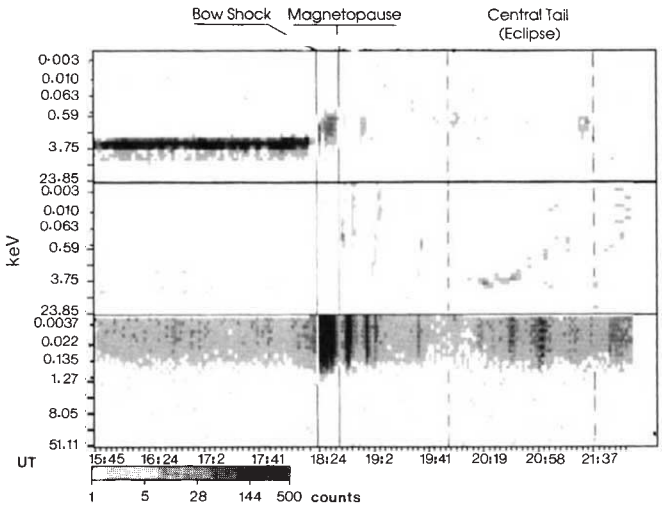


FIG. 1 Energy–time spectrogram of H^+, O^+ and electrons (top, middle and bottom, respectively) flowing away from the Sun. These data were taken during the first orbit of the Phobos 2 spacecraft around Mars, 1–2 February 1989.

The plasma escape from Mars in the tail boundary layer has been calculated[2] to be $\sim 10^{24}$ s^{-1}, corresponding to about $10\,g\,s^{-1}$ of oxygen. This is, however, too low to explain the dehydration of Mars.

The ASPERA (automatic space plasma experiment with a rotating analyser) ion-composition experiment on board Phobos 2 has made it possible to determine more accurately the ion outflow from the upper ionosphere of Mars. As expected, the ion escape is consistent with the cold-plasma composition in the topside ionosphere inferred[4] from the US Viking 1 and 2 spacecraft, that is, oxygen ions dominate. Moreover, the ASPERA data from the first four Phobos 2 orbits around Mars indicate that the ionospheric loss rate in the martian magnetotail is high enough to explain the dehydration of Mars over its lifetime.

The instrument ASPERA is a plasma-composition experiment that uses a scanner platform to provide nearly complete (that is, three-dimensional) coverage of the unit sphere of the three-axis-stabilized spacecraft. The ASPERA measures the composition, energy and angular distribution of ions of charge e with energies $0.5\,eV\,e^{-1}$ - $24\,keV\,e^{-1}$ and electrons with energies 1 eV - 50 keV. The plasma analyser comprises two spectrometer systems with a 360° field of view lying in a plane perpendicular to the plane of rotation (±90°). The 360° field of view is divided into ten sectors for ions and six sectors for electrons, all sectors containing individual sensor elements. The unit sphere can be covered by a 180° turn of the scanner platform.

Figure 1 shows energy–time spectrogram plots for electrons and H^+ and O^+ ions observed during the first two orbits around Mars. The panels correspond to data (~ 4-min running means, ~ 50-ms sampling time) sampled in the sunward sectors of the ASPERA. The important boundary crossings (such as the bow shock and magnetopause) are marked by vertical bars in the panel. The central tail (eclipse) is contained in the region of dashed lines.

In the first orbit the bow shock was crossed near the subsolar point at an altitude of ~ 1,550 km, and the magnetopause was crossed near the location of the pericentre (~ 860 km) in the dusk sector of Mars. We note from Fig. 1 that the most persistent feature inside the magnetospheric cavity of Mars is the dominance of ionospheric ions (O^+). Near the planet (at ~ 18:30 - 19:00 UT) the outflow of 'cold' ions in the few eV range dominates, whereas the central tail is characterized by accelerated cold beams up to several keV. The existence of accelerated cold beams in the central martian tail was quite unexpected and is one of the remarkable similarities between the martian and terrestrial magnetosphere.

Figure 2 displays the number density and flow velocity (v_x = flow velocity away from the Sun) for H^+ ions and the number density for O^+ and molecular ions during the second orbit. The mass resolution is not sufficient to separately resolve the molecular ion species anticipated in this mass range (such as N_2^+, O_2^+ CO^+ and CO_2^+). It is clear from Fig. 2 that the contribution of molecular ion species in the ionospheric ion escape is indeed significant. At times the molecular ion escape even dominates over the atomic oxygen escape. This implies a considerably higher outflow of molecular ions than that found near the Earth. Figure 2 illustrates the main characteristics of an encounter with the martian magnetosphere—strongly reduced H^+ ion densities and flow velocities and a correspondingly large increase of energized ionospheric-ion densities. The density of O^+ ions reaches up to 6 cm^{-3} even in the distant tail of the martian magnetosphere.

Figure 3 shows a composite diagram of the total outflow (in ions $cm^{-2}\,s^{-1}$) from Mars of O^+ ions along the spacecraft trajectory of the first and second orbits. The O^+ flux in the martian tail generally exceeds 10^6 ions $cm^{-2}\,s^{-1}$. The average O^+ flux during the four first orbits corresponds to $\sim 2\times 10^6$ ions $cm^{-2}\,s^{-1}$ in the distant ($x \sim 10 R_M$) tail with an average width of $\sim 8 R_M$. The O^+ outflow in the central tail was very variable but on the average of similar flux intensity as that in the tail boundary.

Figure 4 summarizes the ionospheric outflow observed during the first four orbits. The diagram illustrates that the escape of ionospheric plasma is predominantly found in the tail boundary of Mars. Cooler and denser plasma flows out in narrow structures

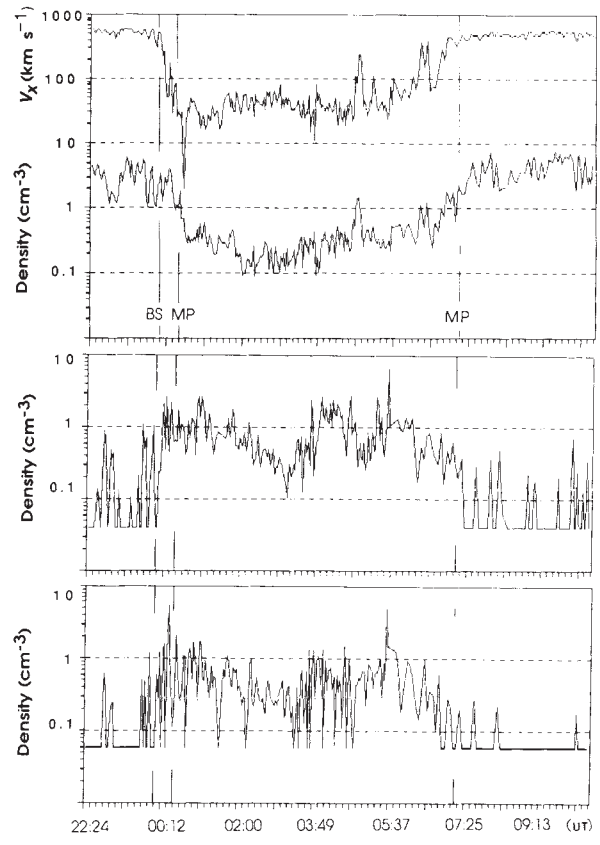


FIG. 2 Moment parameters for O^+ and H^+ ions during the second Phobos 2 orbit around Mars (4–5 February 1989). The top panel gives the flow velocity and the number density of H^+ ions. The middle and lower depict the number density of O^+ and molecular ions (O_2^+, CO_2^+, CO^+), respectively. The inbound bow shock (BS) and magnetopause (MP), and the outbound magnetopause crossings are marked by vertical lines.

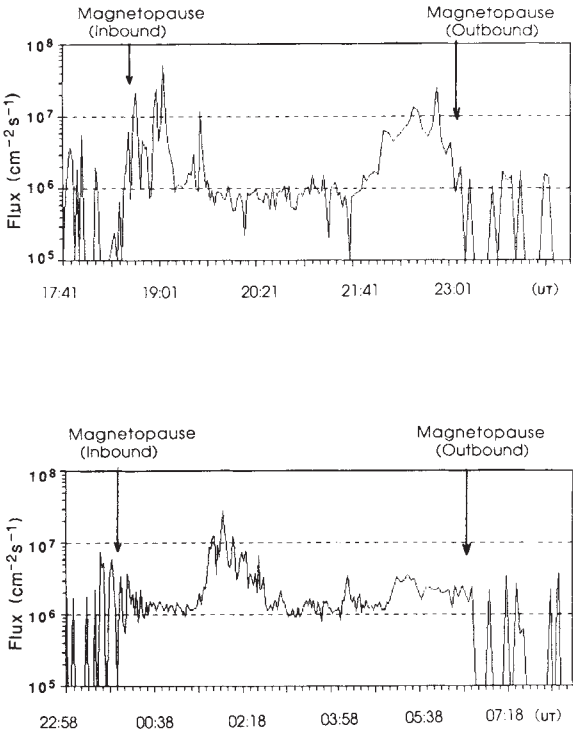


FIG. 3 Composite diagram of the outflow ($n\,v_x$) of ionospheric O^+ ions from Mars during orbits 1 (top) and 2 (bottom).

into predominantly the boundary layer and, subsequently, by the mass-loading process, picks up the boundary-layer flow speed away from the Sun. The central tail region is characterized by a highly variable outflow of O^+ ions which are subject to acceleration processes similar to those found in the terrestrial magnetosphere ('auroral' ion beams).

The rate of ionospheric-ion outflow is indeed very high with fluxes similar to, or exceeding those found in the Earth's geotail. Although the dominant ion constituent of the outflow is oxygen there are strong reasons to believe that the hydrogen-ion escape may be equally significant along the flanks. The hydrogen-ion escape is, however, difficult to distinguish from intruding solar-wind protons.

The atmosphere of Mars, with an atmospheric pressure at ground level of the order of a few millibars, is much thinner and drier than that of the Earth. Furthermore, the lower atmosphere is dominated by CO_2 (95%) and N_2 (2.7%), containing very small amounts of O_2 (0.13%) and H_2O (0.03%). The ionosphere of Mars shows, however, many similarities with the Earth's ionosphere, despite strong differences in the composition of the neutral atmosphere. For instance, the dominating ion species in the ionosphere is oxygen[4] (mainly O^+ at the Earth and O_2^+ at Mars).

The solar-wind interaction with a planet having a conducting ionosphere is associated with a transfer of energy and momentum from the solar wind to the topside ionosphere. This results in a heating and loss of plasma from the planetary ionosphere, and consequently also a loss of atmosphere from the planet. Theoretically, assuming a complete transfer of solar-wind momentum from ions (velocity $v_{sw} \approx 400\ \mathrm{km\,s^{-1}}$) impinging on the topside of Mars to the ionospheric ions until they reach the escape velocity ($v_{esc} \approx 5.0\ \mathrm{km\,s^{-1}}$ on Mars), the escape flux could be 80-times higher than the solar-wind (proton) influx. Thus, through a martian central tail with cross-section $\sim 5\times10^{13}\ \mathrm{m^2}$ a (proton) flux of $\sim 7\times10^{27}$ ions $\mathrm{s^{-1}}$ may escape as a result of the momentum exchange with a solar wind of density $5\ \mathrm{cm^{-3}}$. If the escaping particles are oxygen atoms the maximum escape rate would be 16 times lower ($\sim 4\times10^{26}$ ions $\mathrm{s^{-1}}$). This corresponds to a maximum theoretical escape rate of the ionosphere of $\sim 11\ \mathrm{kg\,s^{-1}}$. Thus, the present solar-wind interaction with the martian upper ionosphere is capable of causing a massive erosion of the atmosphere from Mars.

The average total O^+ ion escape observed in the tail was $\sim 2\times10^{25}$ ion $\mathrm{s^{-1}}$, corresponding to an oxygen loss rate of about $0.5\ \mathrm{kg\,s^{-1}}$. Taking into account the observed molecular-ion escape (such as O_2^+, CO_2^+) and the H^+ ion escape implies that the total escape of volatiles from Mars because of the solar-wind interaction is at least $1\ \mathrm{kg\,s^{-1}}$.

The first plasma composition results from the martian magnetosphere obtained by the ASPERA experiment on board the Soviet Phobos spacecraft show that the escape of plasma from the ionosphere has characteristics similar to those found near the Earth. In the central tail upward-escaping ions are accelerated up to keV energies. The massive escape of ionospheric plasma, however, is contained in the martian tail boundary where the Mars/solar-wind interaction is more cometary.

In view of the inventory of volatiles (in both the atmosphere and the hydrosphere) at the Earth, the terrestrial net loss of $\sim 2\ \mathrm{kg\,s^{-1}}$ (2×10^{26} ions $\mathrm{s^{-1}}$, see ref. 5, for example) from the upper ionosphere through the magnetotail is of minor importance. At this rate, it will take at least 10^9 yr to evacuate the present atmospheric oxygen content of the Earth. Because the mass of Mars is ten times smaller, a similar outflow for Mars is, however, quite significant. At the present rate, it will take $< 10^8$ yr to evacuate the oxygen content of a ~ 7 mbar martian atmosphere (chemically bound in CO_2). The present loss rate of oxygen is equivalent to the loss of a layer of 1 m of water on Mars in 4.5×10^9 yr. Because the major ionospheric loss is associated with the solar-wind interaction in the dayside ionosphere, a higher atmospheric pressure on Mars would lead to a more extended exosphere and ionosphere and a correspondingly higher escape rate. As noted previously, an escape rate that is ten times higher than the measured one is theoretically possible. Thus, if Mars also lacked a significant magnetic field in the past, it implies that the planet has been exposed to a continuous atmospheric erosion through the solar-wind interaction. The lack of sufficiently strong intrinsic magnetic fields may be the reason why both Venus and Mars lost most of the water collected during the accretion process. □

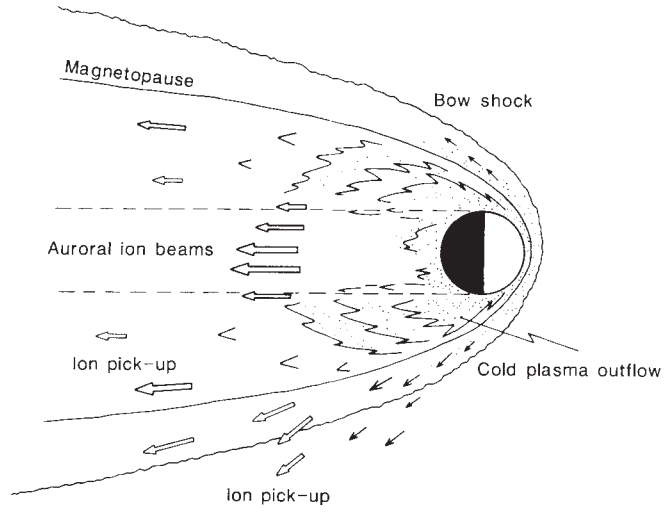


FIG. 4 Representation of the ionospheric ion outflow from Mars. The diagram illustrates the properties of the main ionospheric ion-escape regions, the tail boundary and the central tail.

Received 7 June; accepted 17 August 1989.

1. Gringauz, K. I., Bezrukikh, V. V., Verigin, M. I. & Rezimnov, A. P. *J. geophys. Res.* **81**, 3349–3352 (1976).
2. Vaisberg, O. L. in *Physics of Solar Planetary Environment* Vol. 2 (ed. Williams, D. J.) 8450 (American Geophysical Union, Washington, DC, 1976).
3. McElroy, M. B., Kong, T. Y. & Yung, Y. L. *J. geophys. Res.* **82**, 4379–4388 (1977).
4. Hanson, W. B., Sanatani, S. & Zuccaro, D. R. *J. geophys. Res.* **82**, 4351 (1977).
5. Chappell, C. R., Moore, T. E. & Waite, J. H. Jr *J. geophys. Res.* **92**, 5896 (1987).

ACKNOWLEDGEMENTS. The ASPERA experiment is a joint Swedish, Soviet and Finnish venture supported by grants from the Swedish Board for Space activities and the Academy of Finland. We acknowledge the contributions and technical support provided by the Space Research Institute of the Soviet Academy of Sciences.

Ions of martian origin and plasma sheet in the martian magnetosphere: initial results of the TAUS experiment

H. Rosenbauer*, N. Shutte†, I. Apáthy‡, A. Galeev†, K. Gringauz†, H. Grünwaldt*, P. Hemmerich*, K. Jockers*, P. Király‡, G. Kotova†, S. Livi*, E. Marsch*, A. Richter*, W. Riedler§, T. Remizov†, R. Schwenn*, K. Schwingenschuh§, M. Steller§, K. Szegö‡, M. Verigin† & M. Witte*

* Max-Planck-Institut für Aeronomie, D-3411 Katlenburg-Lindau, FRG
† Space Research Institute, Profsoyuznaya 84/32, 117810 Moscow GSP-7, USSR
‡ Central Research Institute for Physics, H-1525 Budapest, Hungary
§ Institute for Space Research, Austrian Academy of Sciences, A-8010 Graz, Austria

UNLIKE plasma instruments used on previous space missions to Mars, the TAUS instrument on Phobos 2 was designed so that the energy per charge and angular spectra of three species of ions could be measured separately. These species were H^+ and He^{2+} characteristic of the solar wind, and 'heavy ions' collected in one integral channel covering the mass per charge (M/q) range 3 to ∞, which we anticipated to find predominantly in the near-martian regime. In all spacecraft orbits around Mars we found a sharp boundary, separating the shocked solar wind from the martian magnetosphere which was characterized by the absence of solar-wind-like plasma. As the plasma inside the magnetosphere, and particularly in the tail, was dominated by heavy ions with number densities orders of magnitude higher than found in the solar wind, we assumed it was mainly of martian origin. Typically, heavy ions of low tailward flow velocity were seen near the boundary of the magnetotail, whereas high-speed tailward plasma flows of such ions were detected deeper inside the tail, a region not investigated before. Near the centre of the martian magnetotail a plasma regime, comparable to the terrestrial as well as the venusian[1] plasma sheet, was detected, characterized by highly supersonic tailward streams of heavy ions. The flux of planetary ions leaving Mars through its magnetotail is tentatively estimated to be of the order of a few times 10^{25} s^{-1}. Such loss rates would be significant for the dissipation of the martian atmosphere on cosmological timescales.

The instruments previously used for investigating the martian plasma environment were either plasma cups or curved electrostatic analysers by which different types of ions generally cannot be distinguished. To resolve this deficiency, the TAUS spectrometer, which uses electric and magnetic fields for separating ions according to their M/q, was specially designed for investigation of the solar wind and its interaction with Mars. The instrument features a field of view of ~40°×40° centred on the nominal direction towards the Sun and divided into 8×8 channels for angular resolution. Its energy per charge (E/q) range of ~30 V–6 kV is subdivided into 32 channels. In the H^+ and He^{2+} channels, three-dimensional (3D) spectra, resolved in energy, azimuth and elevation can be measured whereas the spectra in the heavy-ion channel can only be resolved in E/q and elevation. The sensitivity of the instrument is nearly constant over its entire field of view but depends on the ion speed. The one-count detection limit for density, for example for a cold O^+ plasma with a speed of 100 km s^{-1}, would be $\sim 1\times10^{-3}$ cm^{-3}. As a data rate of only ~10 bits s^{-1} was available for most of the time, the instrument's high resolution in velocity space and time could not be fully exploited; instead the information had to be compressed on board.

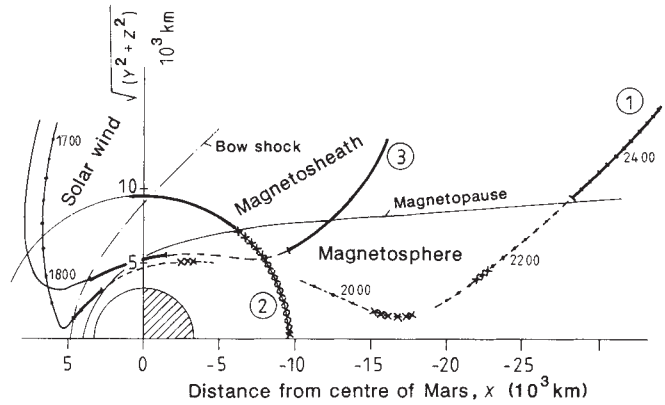


FIG. 1 The orbits of Phobos 2 on 1 February, ①; and on 2 March, ②; and of Mars-5 on 13 February 1974, ③; are given in cylindrical, Mars-centred solar-ecliptic coordinates. Approximate positions of the most important boundaries, the bow shock and the magnetopause, as derived from TAUS data, are shown, too. Along orbit ①, time tags are given for easy reference to Fig. 2. Those parts of the trajectories where the spacecraft passed through solar wind are indicated by thin full lines, shocked solar wind is marked by enhanced full lines, and the passage through the martian magnetosphere is indicated by dashed lines. Crosses on the dashed lines show the positions where supersonically tailward streaming plasma was found. Gaps in the orbital lines indicate that no data were received.

As the first orbit of the Phobos 2 spacecraft around Mars, on 1–2 February, provided a good survey of the martian plasma environment, we will demonstrate our initial results using mainly these data. The Phobos spacecraft trajectory for this period is shown in Fig. 1 in Mars-centred solar-ecliptic cylindrical coordinates.

In Fig. 2*a*, we present 4-min averages of proton count rates as a function of particle energy, with time progressing from bottom to top. The lower spectra show that the solar wind with a speed of ~750 km s^{-1} and a temperature of $\sim 150\times10^3$ K was steady. The sudden decrease of mean energy and broadening of the spectra at ~18:25 UT is the signature of the martian bow shock, which indicates the first major interaction of the solar wind with the 'obstacle' Mars.

Between the bow shock crossing and the transition through the magnetopause at ~18:37 UT, the observed spectra were broad and fluctuating as is typical of shocked solar wind in the magnetosheath. On crossing the magnetopause, the proton fluxes fell below the sensitivity threshold of our instrument and remained at this level everywhere in the martian magnetosphere up to ~23:10 UT, when broad turbulent proton spectra indicate that the spacecraft was in the magnetosheath again.

These characteristic variations of proton spectra in near-martian space have been observed before, for example by instruments on spacecraft Mars 5 (ref. 2). Bow-shock positions determined by Phobos instruments are also close to those obtained in previous martian missions.

The most interesting new result of the Phobos plasma

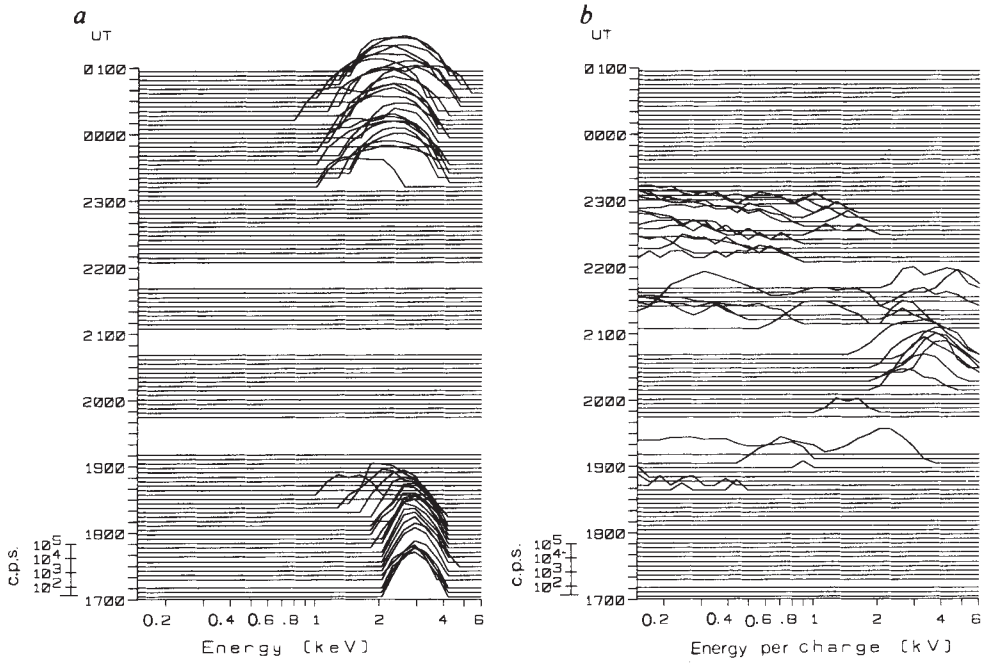


FIG. 2 Four-minute averages of energy-per-charge spectra in terms of count rates of protons (*a*) and heavy ions (*b*) as a function of time including the passage through the martian magnetosphere of the Phobos 2 spacecraft during its first martian orbit on 1–2 February are presented. The foot-lines of the spectra correspond to the one-count level or 40 counts per second. Average background count rates have been subtracted, and cross talk from the proton channel to the heavy-ion channel removed. Blanks indicate data gaps; c.p.s., counts per second.

measurements is the discovery of strong fluxes of heavy ions in the martian magnetosphere. Figures 2*b* and 3*b* show examples of E/q spectra in terms of count rates recorded by the TAUS heavy-ion channel while the spacecraft was crossing the martian magnetosphere. Figure 2*b* shows that heavy ions with energies per charge $\leqslant 1$ kV were observed on 1 February immediately after the entry of the spacecraft into the magnetosphere. Although the maximum ion energy per charge is rather high, the plasma may be nearly stagnant because the maximum of the distribution is found in the lowest E/q channels. By comparison with the opposite side of the magnetotail (22:00–23:00 UT) and with other orbits (see Fig. 3*b*) it appears that this type of heavy-ion plasma is typical of the boundary region of the martian magnetosphere traversed by the Phobos spacecraft. However, in a few cases, supersonically tailward-streaming plasma was observed just inward of the magnetopause as well.

As the spacecraft proceeds further into the magnetotail, a high-energy peak develops in the spectra that eventually, near the centre of the tail, remains as the sole feature of the distributions. This means that the plasma has a well-defined, highly supersonic tailward flow speed in that region. As this type of fast tailward streaming plasma was also found in other orbits (see Fig. 3*b*) always near to the region where the tailward component of the magnetic field reverses sign and the total field strength is close to a minimum, we conclude that this is the region corresponding to the plasma sheet in the geomagnetic tail. However, whereas the Earth's plasma sheet is normally populated by solar-wind-type plasma, and heavy ions of ionospheric origin are only found in appreciable amounts during substorms, the martian plasma sheet seems to consist, at least in its central region, of supersonically tailward-streaming heavy-ion plasma most of the time. Protons were seen only in a few cases. Inspection of all of the tail crossings shows that the position of the martian plasma sheet is considerably more vari-

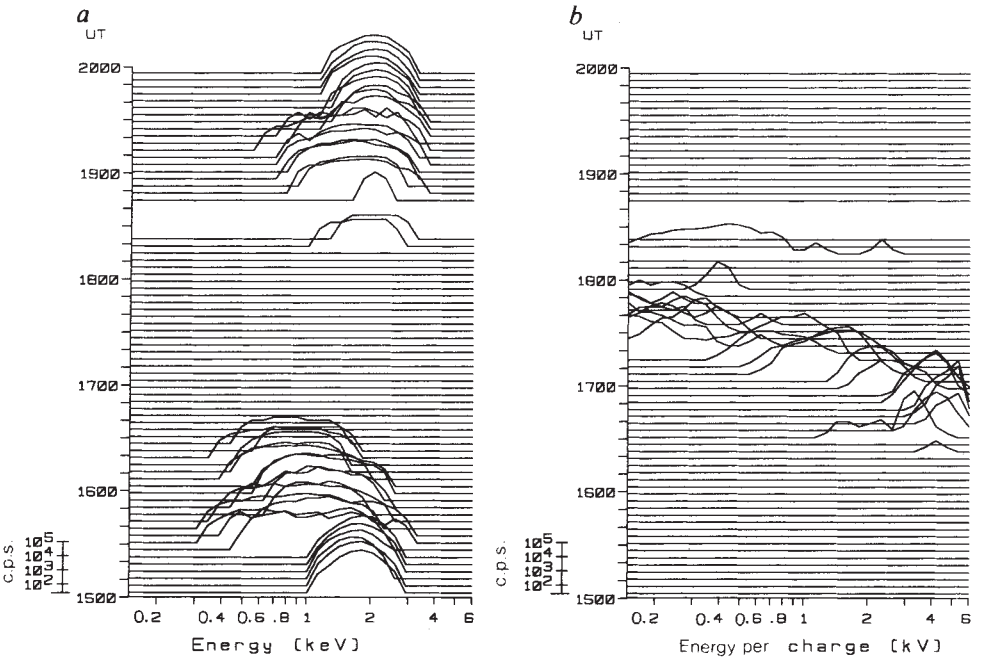


FIG. 3 Energy-per-charge spectra are shown for the circular orbit on 2 March 1989 (see curve ② in Fig. 1) in the same format as in Fig. 2. In the proton data (*a*), the inbound and outbound bow-shock crossings at 15:28 UT and 19:22 UT respectively are clearly visible. Although the maximum flow speed of heavy ions (*b*) was found close to the boundary of the magnetotail, its position again coincides with the reversal of the axial component of the magnetic field, typical of the plasma sheet, and the plasma characteristics are similar to those presented in Fig. 2 where the maximum of the flow speed was measured close to the centre of the tail.

able than the Earth's; also, the density and speed of the plasma found are different from orbit to orbit. This is probably due to the large amount of 'induced' magnetic field in the martian magnetosphere introducing most of the interplanetary magnetic field fluctuations into the tail.

The ion-velocity distributions found in the martian plasma sheet are also worthy of note. As an example, two two-dimensional spectra measured in the plasma sheet region during the second and the fourth elliptical orbit, at a distance of $\sim 1.5 \times 10^4$ km from the centre of Mars, are shown in Fig. 4. Although the mean tailward velocities of the plasma sheet ions were vastly different, the distributions are similar: the transverse (in the instrument's frame of reference) velocity spread or temperature is much higher than the longitudinal one and the resulting anisotropy of ~ 7 is probably one of the largest ever observed in a plasma anywhere. These velocity distributions are regarded as a very important diagnostic tool for the identification of the ion-acceleration processes at work.

Another interesting aspect of these E/q distributions is that they are only singly peaked. This strongly indicates that the observed plasma consists essentially of one ion species only. The high density of these heavy ions, which exceeds that of the heavy ions ($M/q \geq 3$) in the solar wind by about three orders of magnitude, leaves us no choice but to assume that they are of martian origin. The known composition of the upper martian ionosphere[3] from where they have probably been removed makes it highly probable that they are O^+.

It is of interest to estimate the total flux of heavy ions lost by the planet through its magnetotail. As the largest ion fluxes are observed in the plasma sheet, and we do not yet know much about plasma-sheet structure and variation with time, this initial evaluation must be regarded as tentative. We assume that we do not miss a major part of the ion flux as a result of the limited field of view of our instrument and take the width of the plasma sheet to be one martian diameter, $2R_M$, and to extend over the whole tail diameter of $\sim 4.5R_M$. From the data obtained during the second elliptical orbit the density and velocity of O^+ ions in the plasma sheet can be estimated to be of the order of $1.5\ cm^{-3}$, and $150\ km\ s^{-1}$, respectively. This results in a total tailward flux of O^+ ions of $\sim 2 \times 10^{25}\ s^{-1}$, meaning that Mars would be deprived of all of its presently existing atmosphere after $\sim 10^9$ years, if the mass loss rate did not change with time and no other loss or gain processes existed. □

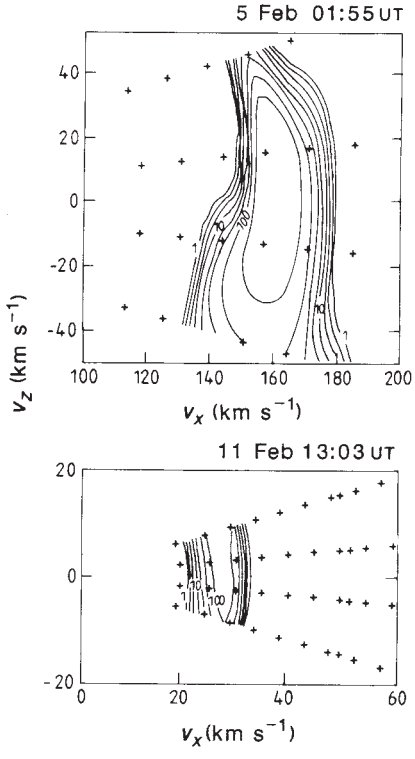


FIG. 4 Contour plots of count rates (to first order representing two-dimensional velocity distributions) obtained in the heavy-ion channel during the second (top) and fourth (bottom) orbit near the plasma sheet. Lines are spaced logarithmically by a factor of $\sqrt[3]{10}$ between adjacent contours. Crosses indicate the centres of the measurement channels. Conversion from E/q to velocity was made by assuming that the ions are O^+. V_x is the anti-sunward component of the velocity, and V_z is perpendicular to V_x. (The direction of V_z relative to the ecliptic plane cannot yet be given because of still missing information on the roll angle of the spacecraft.)

Received 20 June; accepted 7 September 1989.

1. Slavin, J. A. *et al. J. geophys. Res.* **94**, 2383–2398 (1989).
2. Gringauz, K. I. *Rev. Geophys. Space Phys.* **14**, 391–402 (1976).
3. Hanson, W. B. *et al. J. geophys. Res.* **82**, 4351–4363 (1977).

Observation of electron and ion fluxes in the vicinity of Mars with the HARP spectrometer

N. M. Shutte*, P. Király†, T. E. Cravens‡, A. V. Dyachkov*, T. I. Gombosi‡, K. I. Gringuaz*, A. F. Nagy‡, W. E. Sharp‡, S. M. Sheronova*, K. Szegö†, T. Szemerey†, I. T. Szűcs†, M. Tátrallyay†, A. Tóth† & M. Verigin*

* Space Research Institute, Profsoyuznaya 84/32, 117810 Moscow GSP-7, USSR
† Central Research Institute for Physics, PO Box 49, H1525 Budapest 114, Hungary
‡ Department of Atmospheric, Oceanic and Space Sciences, Space Physics Research Laboratory, University of Michigan, Ann Arbor, Michigan 48109–2143, USA

THE highly elliptical orbits of the Phobos 2 spacecraft, early in February 1989, proved particularly useful for plasma and field investigations of the martian environment. The low-altitude (~860 km) pericentres and the deep penetration into the magnetotail provided excellent opportunities to explore new and important regions. Here we present preliminary results of electron and ion measurements in the vicinity of Mars with the hyperbolic analyser in the retarding potential mode (HARP). HARP is a differential electrostatic analyser, simultaneously covering eight directions arranged in a fan-shaped geometry, in the anti-solar hemisphere. The angular resolution is ~20°, the energy resolution ~10%. During the first two elliptical orbits, to be discussed here, electrons from 3.4 to 550 eV and ions from 0.25 to 550 eV were measured in 25 and 50 logarithmic energy steps, respectively. The energy distribution of electrons in the magnetosheath was found to be generally characterized by two distinct peaks. A fairly hot electron component was discovered in the plasma sheet of the areo-magnetic tail.

The HARP differential electrostatic analyser allowed the measurement of charged particles within an energy (or, more precisely, energy per charge) range of ~0.2–800 eV. The energy steps of the instrument were adjustable, that is, a more detailed study of some selected energy intervals was also possible by telecommand. Particles were measured simultaneously in eight sectors (20° × 10° each), symmetrically arranged relative to the anti-solar axis. The actual instrument package on the Phobos 2 spacecraft consisted of two independent and identical sensors. The sensors were mounted to point at 90° from each other. Each unit had four independent anodes. The combined directional

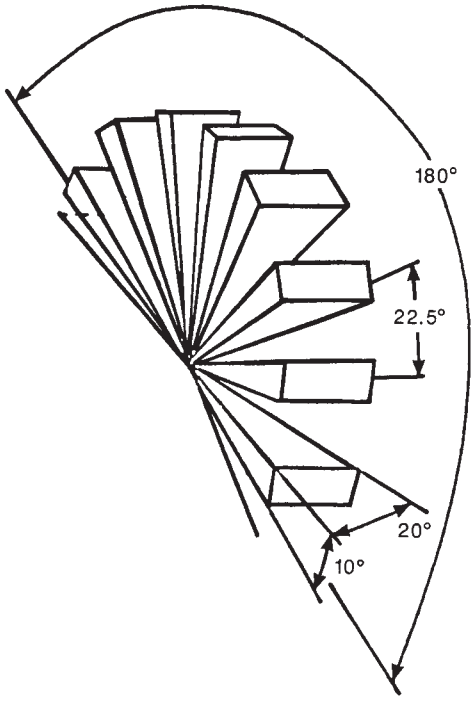


FIG. 1 Directional coverage of the HARP device (centred on the anti-solar direction).

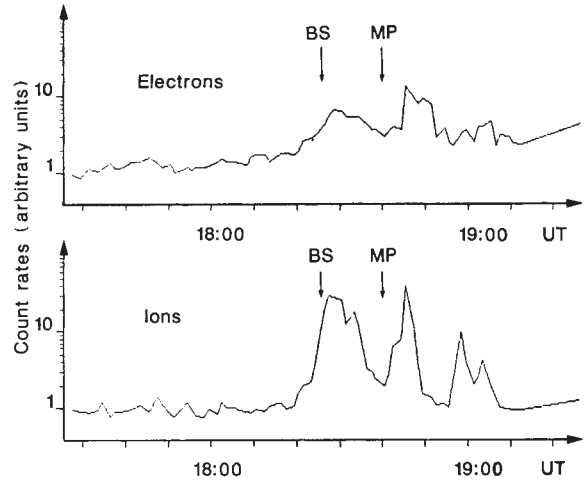


FIG. 2 Electron and ion total count rates as a function of time before and during Mars encounter on the first elliptical orbit of the Phobos spacecraft. BS, bow shock; MP, magnetopause.

coverage is shown in Fig. 1. In the normal, three-axis-stabilized mode of the spacecraft, the viewing directions were in a plane perpendicular to the ecliptic. During the two orbits discussed here there was a slow rotation around the symmetry axis, and the precise attitudes as a function of time are not yet available. The Phobos HARP device[1,2] was specifically developed for this project. Earlier versions[3] were used in terrestrial ionospheric research.

The instrument's energy coverage during the first two elliptical orbits was as specified above. The integration period at each energy step was 1 s. The electron and ion measurements were made sequentially, which means that a complete electron/ion energy and angular distribution was obtained in 75 s. During the high-bit-rate telemetry regime covering the Mars encounters, practically full telemetry coverage was available.

The results of the electron measurements are highlighted here. Results of the ion measurements are only used for comparison in Fig. 2, where total count rates of electrons and ions (summed over energy and direction) are plotted as a function of time during the first elliptical orbit as the spacecraft approached the planet. The location of the bow shock and planetopause (magnetopause) depends on how these boundaries are defined in terms of the measured parameters. For example, the bow-shock crossing as indicated by the appearance of high-energy electron fluxes differed by ~5 min from the bow-shock position defined by the magnetic-field experimenters as the location of an observable jump in the magnitude of the magnetic field. Further and more careful intercomparisons and self-consistent definitions will be needed in the future to resolve these issues. In Fig. 2 we indicate the bow shock and planetopause/magnetopause crossing locations chosen by the magnetic field experimenters[4]; the increase of both electron and ion fluxes in the vicinity of the bow shock is clearly visible.

Spectral information provides a more detailed description of the dynamics of electron populations. As an example, the change of the spectral character of the measured electron flux in one directional channel (56° from the symmetry axis) is given in Fig.

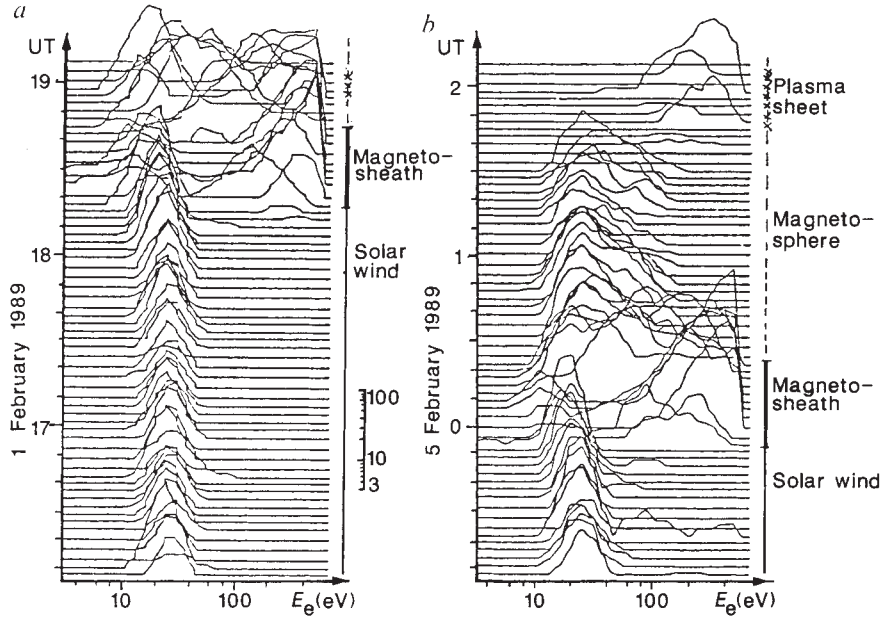


FIG. 3 Variation in time of the energy spectra measured by the HARP device during *a* the first and *b* the second Mars encounter of the Phobos spacecraft.

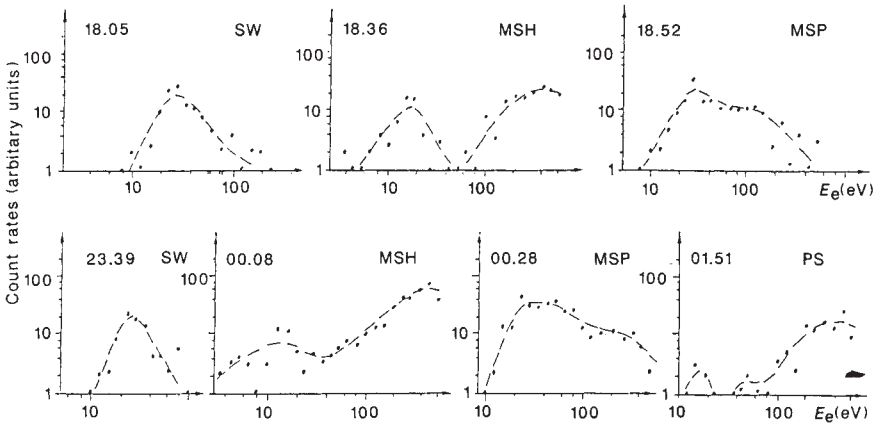


FIG. 4 Characteristic energy spectra (count rates as a function of energy) on the first two highly elliptic orbits. *a*, 1 February 1989; *b*, 4–5 February 1989. SW, solar wind; MSH, magnetosheath; MSP, magnetosphere; PS, plasma sheet.

3 for the first two elliptical orbits. (The 'spectra' represented here are really count rates proportional to energy flux, as usually measured by electrostatic analysers. The baselines represent 3 counts s^{-1} even when the measured count rate was lower.) For the first orbit, the sharp change in the spectral shape of the electron flux indicates a bow-shock position at 18:20 UT on 1 February, corresponding to a distance of ~1,500 km from the surface of the planet and a solar zenith angle (SZA) of 10°. For the second orbit, the spectral change, indicating the bow shock, occurred at 23:55 UT on 4 February, corresponding to a distance from the surface of ~1,700 km and a SZA of 11°. The periapsis of the orbits at a distance of ~860 km from the surface was at 18:39:37 UT on 1 February and 0:15:47 UT on 5 February.

The bow-shock crossing times determined from the HARP electron spectra are in reasonable agreement with those obtained by the MAGMA magnetometer[4] and the TAUS ion spectrometer[5]. The changes in the total flux and energy distribution of the measured spectra inside the bow shock are clear indications of changing plasma regimes. The solar-wind interaction processes at Mars are believed to be different from those at Earth and Venus, therefore the accepted terminologies used to delineate different regimes at these planets may not be quite appropriate to Mars. However, it is premature to coin new boundary definitions; therefore, we use here the terminology accepted for the terrestrial magnetosphere to label the different regions in Fig. 3, as delineated by the observed spectra. The data presented in Fig. 3 indicate that the electron spectra observed in the tail region of the planet are similar to those in the region just inside the bow shock. The energy spectra clearly show that electrons with energies ≫100 eV are present inside the bow shock in the magnetosheath.

Figure 4 presents typical spectra measured in various regions of the magnetosphere. (These spectra represent uncorrected count rates, so even single counts are included, in contrast to the baselines of Fig. 3.) In the magnetosheath the electron energy distribution is strongly non-maxwellian, often showing a distinct double peak; significant fluxes are present at energies beyond the energy limit of HARP. In the inner magnetosphere the electron fluxes were found to be generally more isotropic than in the magnetosheath. The corresponding energy distributions had a rather broad maximum, extending from ~20 to 150 eV. Most spectra are also characterized by a double-peak structure. In the shadow of the planet, the energy distributions of electrons, as mentioned above, are similar to those observed in the magnetosheath. This region, in the tail, is referred to as the plasma sheet. Before reaching the tail, the intensity of the electron fluxes dropped and showed strong fluctuations. Our identification of the plasmaphysical regions, as indicated in Figs 3 and 4, was supported by the TAUS ion measurements on the Phobos spacecraft[5].

Summarizing, the measurements of electron spectra along the orbit show sharp changes in both flux levels and in spectral shapes as the spacecraft crosses the bow shock, magnetopause and plasma sheet. The similarity of the observed electron spectra in the magnetosheath and the plasma sheet is likely to be an indication that the plasma sources and/or acceleration processes are similar in these regions. The plasma sheet with accelerated electrons in the inner magnetotail of Mars was detected for the first time with the HARP experiment on board the Phobos spacecraft.

The physical mechanisms responsible for the acceleration of the electrons inside the bow shock and the creation of double-peaked energy distributions are not known at this time. However, further analysis of the HARP data and careful comparisons with the observations of other field and particle experiments carried by the Phobos spacecraft, as well as with previous observations[6], will help to advance our understanding of Mars's plasma environment. □

Received 19 June; accepted 29 August 1989.

1. Szücs, I. T. *et al. Nucl. Instrum. Meth. Phys. Res.* A (in the press).
2. Király, P. *et al.* in *Instrum. Meth. Space Sci. Res., Nauka* (in the press).
3. Shyn, T. W., Sharp, W. E. & Hays, P. B. *Rev. Sci. Instrum.* **47**, 1005–1015 (1976).
4. Riedler, W. *et al. Nature*, **341**, 604–607 (1989).
5. Rosenbauer, H. *et al. Nature*, **341**, 612–614 (1989).
6. Gringauz, K. I. *Adv. Space Res.* **1**, 5–24 (1981).

Energetic ions in the close environment of Mars and particle shadowing by the planet

V. Afonin*, S. McKenna-Lawlor†, K. Gringauz*, K. Kecskemety‡, E. Keppler§, E. Kirsch§, A. Richter§, D. O'Sullivan||, A. Somogyi‡, A. Thompson||, A. Varga‡ & M. Witte§

* Space Research Institute, Profsoyuznaya 84/32, 117810 Moscow, GSP-7, USSR
† Space Technology Ireland, St. Patrick's College, Maynooth, Ireland
‡ Central Research Institute for Physics, PO Box 49, H-1525 Budapest 114, Hungary
§ Max-Planck-Institut für Aeronomie, Lindau, FRG
|| Dublin Institute for Advanced Studies, Dublin, Ireland

THE twin-telescope particle-detector system, SLED, aboard Phobos 2 recorded flux enhancements in the range 30–350 keV in the same general location in the close environment of Mars, over eight days at ~900 km altitude in three successive elliptical orbits. Here we present possible interpretations of these observations. Energy-related particle shadowing by the body of Mars was also detected, and the data indicate that this effect occurred in <20% of the 114 circular orbits around Mars because of the nutation

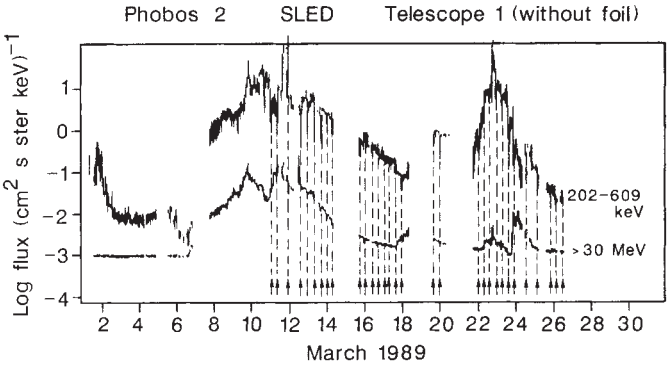


FIG. 1 In the circular orbits around Mars (~6,900 km) in March 1989 the particle fluxes observed by SLED were enhanced because of solar and interplanetary activity. Arrows (bottom) indicate significant depressions in flux resulting from the shadowing by the body of the planet.

of the spacecraft. We discuss the influence of magnetic fields in allowing particles to reach the detector under potentially screened conditions.

The charged-particle detector SLED aboard the three-axis-stabilized Phobos 2 spacecraft was designed to measure the flux of ions and electrons simultaneously in the 30 keV to a few MeV range[1]. The instrument used silicon surface barrier detectors, two in each of the two telescopes viewing in the same direction. One telescope (Te 2) was covered by a thin Al-foil (500μg cm^{-2} Al on Mylar) whereas the other telescope (Te 1) was 'open'. The front detectors of both telescopes were covered by a 15μg cm^{-2} Al-layer. Low-energy protons and electrons could be distinguished[2], as the foil-telescope stopped ions up to 350 keV (with respect to protons) due to the Al-foil and to the Al-layer on the front detector. The geometric factor of each telescope was 0.21 cm^2 ster and the FOV (field of view) axis, with a 40° apex angle, was in the ecliptic plane at 55° to the west of the sunward direction (nominal direction of the interplanetary magnetic field at Mars).

Back detectors were used to discriminate particles which penetrated the front detectors. In addition, each telescope was shielded by 5.6 g cm^{-2} Al and Ta to prevent protons with energies <70 MeV and electrons <10 MeV from reaching the detector systems. Thus, six different energy channels for each telescope could be defined. In the open telescope Te 1, ions and electrons were recorded over the following ranges: channels C1 (30–50 keV), C2 (50–200 keV), C3 (200–600 keV), C4 (0.6–3.2 MeV), C5 (3.2–4.5 MeV), C6 (>30 MeV). In the foil-covered 'electron' telescope Te 2, the energy ranges covered by individual channels were very similar to the above. Channels 6 in both telescopes represent the count rates of the back detectors. In Te 2, protons with energies <350 keV, helium < 1.6 MeV and oxygen ions <8 MeV were stopped in the Al-layers. SLED was the first instrument to approach Mars as closely as 867 km, capable of measuring particles at energies ≥30 keV (refs 3–5).

During the interval from switch-on (25 July 1988) until arrival at Mars (29 January 1989), SLED monitored interplanetary intensity variations with signatures typical of solar minimum conditions (co-rotating interaction regions) under gradual replacement by signatures characterizing solar maximum conditions (produced by transient events such as flares and coronal-mass ejections). Figure 1 provides an example of solar-flare-related particle enhancements recorded by SLED in C3 and C6 of Te 1 when close to Mars in successive circular orbits around the planet during March 1989. These data resemble those observed near the Earth by the satellite GOES 7 (ref. 6).

In addition to observations during the cruise phase, particle measurements were also carried out during four elliptical orbits about Mars, 1–14 February 1989 (pericentre 867 km), and during the following single elliptical and 114 circular orbits (height above Mars 6,000 km). At the time of the first close elliptical orbit, the environment of the planet was greatly disturbed by the presence of particles of solar origin with energies up to a few MeV. Figure 2 (left-hand panels) shows data obtained during the relatively quiet interplanetary conditions characterizing orbit 2 and in more disturbed conditions for orbits 3 and 4. Complementary drawings showing the spacecraft trajectory in polar coordinates (times indicated are in UT) and the position during each orbit of the pericentre (PC), as well as the locations during orbits 2 and 3 of the bow shock and magnetopause, are presented in the right-hand panels. These latter data were provided by the cold plasma (TAUS) and magnetic field (MAGMA) experimenters. Figure 2 shows that, on 5 February 1989, a well-defined enhancement in fluxes increasing by a value of about two orders of magnitude and a duration of 8 min was recorded near pericentre in Te 1, C1 (30–50 keV), when the spacecraft was ~900 km above the planet. A less pronounced peak of about half an order of magnitude was recorded in C2 (50–200 keV). No enhancement was observed in the foil telescope Te 2, so that the increases recorded in Te 1 seem to have been produced by ions in the range 30–200 keV.

In orbit 3 (see Fig. 2) particle fluxes increased in Te 1, channels 1–3, again near pericentre. The duration of this effect was ~26 min. The largest enhancement occurred in the lowest energy range. The peaks of the enhancements in C2 and C3 occurred ~5 min earlier than that in C1. No special response was observed in Te 2.

In orbit 4 there was a telemetry gap close to pericentre but, as shown in Fig. 2, from after about 4 min, the declining phase of a flux enhancement was present in the data of Te 1, C1. Again, Te 2 showed no flux enhancement.

Positive responses in Te 1 and no responses in Te 2 at the times of pericentre passages may suggest pulse pile-up caused by lower-energy particles[7]. For SLED, baseline restoration is applied, the pulse-shaping time constant is 0.3 μs, the pulse length is 1 μs, and the coincidence resolution time is 1 μs. These factors prevent a pulse pile-up up to 10^5 c.p.s. The maximum fluxes observed near pericentre, however, do not exceed 10^3 c.p.s.

The fact that particle increases were recorded by Te 1 in the same general location, <900 km above the planet over ~8 days, therefore suggests the existence of a zone of enhanced radiation inside the martian environment. By identifying the nature and origin of this radiation, SLED could make an important contribution to the solution of an essential, yet unresolved, question concerning the geophysics of Mars, namely whether or not this planet has an intrinsic magnetic field and thus a magnetosphere, like the Earth.

If Mars indeed possesses a magnetosphere, these observed enhancements could represent a zone of trapped ion radiation (30–350 keV) similar to the van Allen belt in the inner part of the Earth's magnetosphere. In the absence of unambiguous data concerning the overall structure of the ambient magnetic field, this possibility is not excluded. Because the gyroradii of these ions are approximately equal to the height of the spacecraft above Mars, however, these particles could only be quasitrapped, as they would be quickly lost because of their interaction with the dense atmospheric layers of the planet. This provides one possible explanation of the observation that, during orbit 3, the position of maximum flux recorded in the channels of Te 1 shifted with time towards lower energies, with the largest flux detected in the lowest-energy channel. On the other hand, the interaction of a martian magnetosphere with the interplanetary magnetized medium would possibly also lead to a local and sporadic merging of planetary with interplanetary magnetic field lines at the front side of the planet and thus, like at Earth, to an acceleration of charged particles. This mechanism could conceivably provide a source for the observed <350 keV ions, which could then propagate along the magnetopause boundary layer from the day to the nightside of the planet, as already

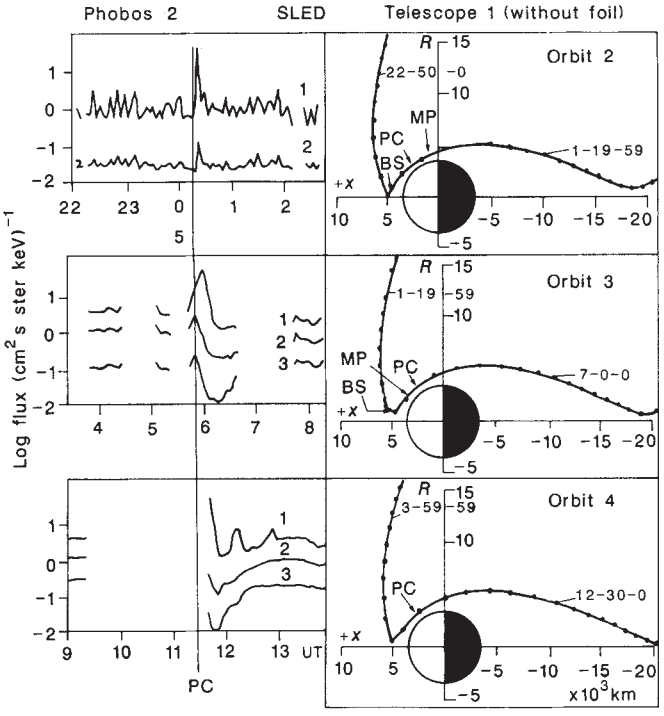


FIG. 2 Particle fluxes (left-hand panels) measured with Te 1 in the highly eccentric orbits 2, 3 and 4 at 4, 8 and 11 February 1989, respectively. In all three orbits clear flux enhancements were observed just after passing pericentre (PC) at ~900 km altitude. The energy bands are: 1, 30–50 keV; 2, 50–200 keV; and 3, 200–600 keV. The right-hand panels show for the same periods the radial distance R (in 1,000 km) of the spacecraft from the x axis (Mars–Sun line) and the positions of the bow shock (BS) and the magnetopause (MP) as measured by the magnetic field and plasma experiments.

observed in the case of the Earth's magnetosphere[8]. It is interesting to note that, at the part of the orbital trajectory concerned, SLED viewed directly along the surface of the magnetopause.

A more general explanation concerning the origin of these observed enhancements could alternatively rest in the fact that the magnetic-field observations indicated a pronounced compression of field lines close to the planet[9]. Such a compression would also increase the density of the ambient plasma and could simultaneously lead to an adiabatic acceleration of suprathermal particles to keV energies inside the martian environment. The acceleration of martian ions, such as oxygen, by the pick-up process in the solar wind can be neglected (provided no additional acceleration by other processes is imposed), as the expected energy would be ~60 keV and therefore below the threshold of the detectors with respect to oxygen.

At present other, more mundane, explanations cannot be ruled out entirely: for example, the observed enhancements could simply be caused by local changes in the angular distribution of particles as the spacecraft traversed different plasma regimes. For directed flux, an intensity enhancement would be observed if the angle between the flow direction and the instrument aperture became favourable. As is well known, the magnetic field changes its direction at the magnetopause quite substantially. Moreover, the possibility that sunlight reflected from the body of the planet (rather than from its thin atmosphere) may have contributed to the observed intensity increases cannot be excluded entirely. As the front detector of Te 1 is only shielded by 15 μg cm^{-2} Al, this detector will be sensitive to sunlight. However, the observed maximum fluxes near pericentre during orbit 3 cannot solely be caused by photons. Because photons will normally not produce a time dispersion in the time-intensity profiles, and the photon fluxes reflected from Mars, when taking the photon fluxes of the Sun at 1 AU (ref. 10), the albedo of Mars of ~10% and the characteristics of SLED into account, can hardly contribute to the observed count rates of ~10^3 c.p.s. and, at the same time, provide responses to C3 (200–600 keV). Yet the effect of sunlight must be investigated in more detail when adequate information concerning the attitude of the spacecraft becomes available. Finally, the possibility that the observed intensity increases are a result of bremsstrahlung[5], produced by low-energy electrons trapped along magnetic field lines, can be excluded, as the electron telescope Te 2 did not detect any electron fluxes.

An additional feature shown by the data of Fig. 2 from orbits 3 and 4, is the depression in fluxes by about one order of magnitude, which is seen particularly clearly in the data of Te 1, C3, C4 and C5. Although, as mentioned above, particle enhancements were not detected in Te 2, the depressions in fluxes recorded by Te 1 were clearly present in the data of C3 and C4 of Te 2. Again, depressions in fluxes recorded during orbit 4 in Te 1, C2, C3, C4 and C5 had well-defined counterparts in Te 2, C3 and C4. These decreases were energy-dependent in both telescopes.

Shadowing[11] by the planet Mars over that part of the orbit where the apertures are directed towards the planet is involved in producing these observed depressions in intensity. It is noted that data taken during 114 circular orbits at an altitude above the planet of 6,000 km indicate that marked decreases in the particle fluxes occurred in <20% of these revolutions, probably because of the nutation of the spacecraft. Successive instances of the occurrence of particle shadowing during the succession of flare-related particle enhancements recorded by SLED when in circular orbit during March 1989 are indicated by arrows in Fig. 1. This sequence seems to indicate the influence of magnetic effects in allowing particles to arrive at the detector during disturbed interplanetary conditions, even in situations where the solid angle of the aperture of the instrument was completely filled by the body of the planet. □

Received 23 June; accepted 25 August 1989.

1. McKenna-Lawlor, S. *et al. Nucl. Instrum. Meth.* (in the press).
2. Anderson, K. A., Lin, R. P., Potter, D. W. & Heetderks, H. D. *IEEE Trans. GE-16*, 153–156 (1978).
3. van Allen, J. A., Frank, L. A., Krimigis, S. M. & Hills, H. K. *Science* **149**, 1228–1233 (1965).
4. O'Gallagher, J. J. & Simpson, J. A. *Science* **149**, 1233–1239 (1965).
5. Vernov, S. N. *et al. Kosmitch. Issled.* **12**, 252–263 (1974).
6. *Solar-Geophysical Data* (No. 536) (NOAA, Boulder, Colorado) 165–180 (1989).
7. Armstrong, T. P., Krimigis, S. M. & Lanzerotti, L. J. *J. geophys. Res.* **80**, 4015–4017 (1975).
8. Richter, A. K., Keppler, E., Axford, W. I. & Denskat, K. U. *J. geophys. Res.* **84**, 1453–1463 (1979).
9. Riedler, W. *et al. Nature* **341**, 604–607 (1989).
10. Feng, W., Ogawa, H. S. & Judge, D. L. *J. geophys. Res.* **94**, 9125–9130 (1989).
11. van Allen, J. A. & Ness, N. F. *J. geophys. Res.* **74**, 71–93 (1969).

ACKNOWLEDGEMENTS. We thank all those who contributed to the SLED experiment and its data evaluation at the six participating institutions, including in particular the Irish technical team M. Martin, B. Jordan, D. Gleeson and N. Murphy, together with K. Fischer of MPAe. We also thank the Irish National Board for Science and Technology (now EOLAS), Space Technology Ireland and the Bundesministerium für Forschung und Technologie for financial support.

T-cell receptor cross-linking transiently stimulates adhesiveness through LFA-1

Michael L. Dustin & Timothy A. Springer

Center for Blood Research, 800 Huntington Avenue, Boston; and Committee on Cell and Developmental Biology and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

Effective interaction between T cells and their targets requires that recognition of specific antigen be coordinated with increased cell–cell adhesion. We show that antigen–receptor cross-linking increases the strength of the adhesion mechanism between lymphocyte function-associated molecule-1 (LFA-1) and intercellular adhesion molecules (ICAMs), with intracellular signals transmitted from the T-cell antigen receptor to the LFA-1 adhesion molecule. The increase in avidity is rapid and transient, providing a dynamic mechanism for antigen-specific regulation of lymphocyte adhesion and de-adhesion.

T-CELL immune recognition requires adhesion receptors as well as the T-cell receptor (TCR), as shown by the ability of monoclonal antibodies directed against these structures to inhibit antigen-specific responses[1]. T-cell adhesion receptors include LFA-1, which binds to its counter receptors ICAM-1 and ICAM-2 (refs 2–4), and CD2, which binds to its counter receptor LFA-3 (ref. 5). LFA-1 is a member of the integrin family, a group of extracellular matrix and cell-adhesion receptors which integrate the extracellular environment with the cytoskeleton[6–8]. The other molecules, ICAM-1, ICAM-2, CD2 and LFA-3 are members of the immunoglobulin superfamily[9]. One unresolved question is how the TCR and adhesion receptors cooperate to balance the competing needs for sensitive antigen recognition and stable cell–cell adhesion. At one extreme, requiring TCR interaction with antigen to contribute the antigen-specific adhesive component, would need a large number of receptor–ligand interactions, lowering the sensitivity of antigen recognition. At the other extreme, strong adhesion of T cells to other cells regardless of antigen expression would lower the efficiency of immune surveillance, as T cells would spend long periods of time in non-productive interactions.

Early studies demonstrated that cytolytic T lymphocytes (CTL) elicited *in vivo* adhered to target cells bearing antigen but not to cells lacking antigen, implying that adhesion was antigen-specific and, by inference, solely due to the TCR[10,11]. With the discovery of adhesion receptors it was initially proposed that these molecules contributed in an additive manner to TCR-specific interactions, so that antigen-specific adhesion could be maintained[12,13], or that they might be involved in an adhesion strengthening step triggered by the TCR[1,14,15]. No evidence for TCR regulation was presented, however, and subsequent studies with CTL lines and clones maintained *in vitro* showed that adhesion was virtually equivalent with both antigen-positive and antigen-negative targets[16,17]. It was proposed that adhesion-molecule interactions preceded TCR binding to antigen, and that the TCR was important only for triggering T-cell effector function[16]; this theory has been widely accepted[18].

Interactions between T lymphocytes and antigen-bearing cells must be reversible. This can be inferred from the existence of circulating T lymphocytes that have previously encountered antigen (memory T lymphocytes) and from studies on CTL. CTL can be observed engaging in repeated cycles of adhesion to target cells, lethal hit delivery and de-adhesion, with a cycle time as short as 15 min[11,14,19]. The mechanism for regulating adhesion and de-adhesion cycles is completely unknown.

We present evidence that LFA-1 function is regulated by phorbol esters and the TCR. Activation of lymphocytes by the TCR rapidly converts LFA-1 to a high-avidity state. TCR engagement thus triggers an adhesion amplification mechanism, allowing antigen-specific adhesion to be driven by a metabolic energy-dependent increase in LFA-1 avidity. The TCR and LFA-1 are coupled by intracellular signalling pathways, as shown by inhibition with dibutyryl cyclic AMP, agents that increase cytosolic cAMP, and the protein kinase inhibitor, staurosporine. The high avidity state of LFA-1 is transient; it peaks 5 to 10 min after TCR stimulation and returns to the low-avidity state by 30 min–2 h, providing a mechanism for de-adhesion.

TABLE 1 Blocking of CD3-stimulated adhesion with monoclonal antibodies

Monoclonal antibody against:	OKT3	OKT3 + anti-IgG2a
	^{51}Cr-labelled resting T-cell binding	
Negative control (X63)	5.0 ± 0.4	56.0 ± 0.1
LFA-1 α (TS1/22)	1.0 ± 0.2	1.0 ± 0.2
LFA-1 β (TS1/18)	0.5 ± 0.1	0.6 ± 0.03
LFA-1 α (TS2/4)	6.8 ± 0.6	57.3 ± 5.5
ICAM-1 (RR1/1)	2.0 ± 0.2	1.5 ± 0.2
CD2 (TS2/18)	5.7 ± 0.3	61.2 ± 5.8

Resting T cells were pretreated with OKT3 (IgG2a; 1:10 culture supernatant) as described in Fig. 1 and the indicated antibody, except RR1/1 which was used to pretreat the plates, for 30 min at 4 °C. All antibodies were washed out. Cells were added to wells with or without 2 μg ml^{-1} horse anti-mouse IgG2a. All antibodies except OKT3 are IgG1 so only the anti-CD3 antibody was cross-linked. ICAM-1 density was 1,000 sites μm^{-2} by ^{125}I-labelled RR1/1 binding. Methods were otherwise identical to Fig. 1. Monoclonal antibodies, as described in Fig. 1. Data shown are representative of two experiments.

Phorbol esters and LFA-1/ICAM-1 adhesion

The first experimental evidence for active regulation of the LFA-1/ICAM-1 adhesion mechanism was the ability of phorbol esters to stimulate LFA-1-and ICAM-1-mediated leukocyte homotypic adhesion[15,20,21]. Adhesion is stimulated rapidly, within one hour, and is not accompanied by any change in cell-surface density of LFA-1 or ICAM-1 (refs 15, 20, 21). It was shown subsequently that LFA-1-mediated adhesion of murine T-cell clones to antigen-negative targets could be enhanced twofold by phorbol 12-myristate-13-acetate (PMA) treatment[22]. Regulation by phorbol esters of LFA-1 avidity, however, could not be distinguished from regulation of ICAM-1 avidity or from a change in some general cellular property such as membrane spreading.

Binding of cells to purified adhesion molecules on inert surfaces is an ideal system to study regulation of adhesion mechanisms. In this situation, only the cell-surface receptor for the purified adhesion molecule is involved in binding, allowing regulation of its avidity (multivalent affinity) to be measured in

isolation from other adhesion mechanisms. The coexistence of ICAMs and LFA-1 on a number of different cell types allowed us to reciprocally assay binding of the same cells to both purified LFA-1 and ICAM-1 on artificial substrates, and so to determine which cell-surface molecule is affected by cellular activation. The hypothesis that PMA stimulates adhesion by increasing the avidity either of cellular LFA-1 or of cellular ICAM-1 predicts that regulation shows sidedness, that is, adhesion to artificial substrates containing purified ICAM-1 or LFA-1 should be differentially affected. Demonstration of sidedness would rule out a general adhesion-promoting effect of phorbol esters, as this should increase adhesion to LFA-1 and ICAM-1 substrates equally. Two distinct LFA-1 counter receptors have been defined, ICAM-1 and ICAM-2. The latter was recently defined by functional complementary DNA cloning[4]. The two immunoglobulin-like domains of ICAM-2 are 35% identical to the two domains closest to the N-terminus of the five immunoglobulin-like domains of ICAM-1. For generality, we have used cells expressing ICAM-1 (JY) as well as those that bind to LFA-1 by an ICAM-1-independent mechanism, which is likely to be through ICAM-2 (SKW3 and resting T cells). Binding of cells to artificial substrates was for six minutes under conditions established to prevent cell-cell aggregation (Fig. 1 legend). Basal binding of JY and SKW3 to ICAM-1 substrates was relatively inefficient over a wide range of ICAM-1 densities (Fig. 1*a*). Adhesion to ICAM-1 was dramatically increased by PMA, however, over a wide range of ICAM-1 densities. By contrast, SKW3 and JY cells adhered to the same extent to LFA-1 substrates in the presence and absence of PMA, although PMA shifted the density of LFA-1 required for half-maximal adhesion by twofold (Fig. 1*b*). Spreading of SKW3 and JY cells on LFA-1 (data not shown) was increased by PMA, which could be the basis of the increased resistance to washing at intermediate LFA-1 densities. Sidedness was even more dramatically illustrated with freshly isolated peripheral blood T lymphocytes. Adhesion to ICAM-1 substrates was almost completely dependent on phorbol ester stimulation (Fig. 1*e*), whereas adhesion to LFA-1 substrates was constitutive and was hardly affected by PMA (Fig. 1*f*). Thus, there is little or no avidity regulation of cell-surface ICAMs and these molecules appear constitutively avid for LFA-1. By contrast, cell-surface LFA-1 is not constitutively avid for ICAM-1, and seems to be converted to a high-avidity state after PMA stimulation.

Temperature effect on LFA-1/ICAM-1 interaction

The LFA-1/ICAM adhesion mechanism is susceptible to inhibition at low temperature (4 °C) in cell-cell binding[14,15] and binding of cells to purified ICAM-1 (ref. 2). This property is relevant

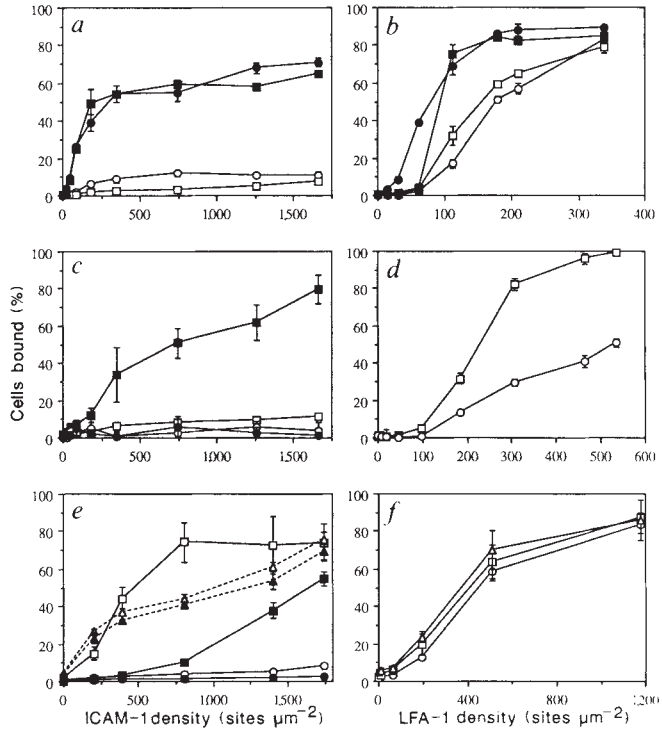


FIG. 1 Effects of activation on binding of cells to purified LFA-1 and ICAM-1. *a*, Binding of JY (circles) and SKW3 (squares) to ICAM-1 adsorbed to plastic without (open) or with (filled) PMA (50 ng ml^{-1}). *b*, same as *a*, binding to LFA-1 adsorbed to plastic. *c*, Binding of JY cells to ICAM-1 at 37 °C (squares) or 4 °C (circles). Cells were treated with PMA (50 ng ml^{-1}) (filled) or with media (open) for 30 min at 37 °C before dispersal and assay. *d*, Binding of JY cells to LFA-1 at 37 °C (squares) or 4 °C (circles). *e*, Binding of resting T cells to ICAM-1; cells pretreated with CD3 monoclonal antibody and then without (open circle) or with (open square) goat anti-mouse IgG, or with PMA (open triangle). Cells treated as above but also pretreated with 2 mM dibutyryl cAMP for 15 min at 24 °C before addition to wells are shown with filled symbols rather than open symbols. *f*, Same symbols as *e*, binding to LFA-1. These data are representative of at least 3 experiments for each group *a–b*; *c–d*, and *e–f*.

METHODS. LFA-1 and ICAM-1 were immunoaffinity purified from JY cell lysates using TS2/4 and RR1/1 sepharose, respectively. ICAM-1 was eluted at pH 12.5 (ref. 2). LFA-1 was eluted at pH 11.5 in the presence of 2 mM $MgCl_2$ (M.L.D. and T.A.S., manuscript in preparation). Purified proteins were at 20–100 μg ml^{-1} in buffered saline with 1% octyglucoside (OG) detergent (+2 mM $MgCl_2$ for LFA-1). Purified LFA-1 and ICAM-1 in 1% OG were adsorbed to polystyrene microtitre plate (Flow, Virginia) wells by addition of 5 μl of the detergent solubilized protein to 45 μl of 25 mM Tris, pH 8.0, 0.15 M NaCl, 2 mM $MgCl_2$ (TSM). After 16 h incubation at 4 °C the plates were incubated for 1 h at room temperature in 1% BSA and TSM, and were then washed with assay media. Site numbers were quantified with ^{125}I-labelled RR1/1 or TS2/4 for ICAM-1 or LFA-1, respectively, at 10 μCi μg^{-1} at a final antibody concentration of 10 μg ml^{-1}. JY was pretreated with TS1/22 (anti-LFA-1 α) for binding to LFA-1, and RR1/1 (anti-ICAM-1) for binding to ICAM-1, and free mAb was completely washed out before the assay. This was the best way to prevent PMA-stimulated JY cell aggregation and did not effect binding to the adsorbed proteins. Similar results were obtained without the antibody pretreatment, but PMA-stimulated cells contained many small aggregates (2–10 cells) especially on ICAM-1 substrates. Other cells gave no homotypic aggregation within the 6-min assay period. Resting T cells were isolated from whole blood by plastic adherence and nylon wool filtration and were 97% CD2$^+$, 93% CD3$^+$ and 96% CD6$^+$. Resting T cells were used within 24 h of drawing blood. T cells were pretreated with anti-CD3 monoclonal antibody (Leu4 ascites at 1:500 or OKT3 culture supernatant at 1:10 as indicated) for 30 min at 4 °C and then washed 3 times at 4 °C. In all assays ^{51}Cr labelled cells (1–2 × 10^4 JY or SKW3, or 5 × 10^4 resting T cells) were added to wells which contained 50 ng ml^{-1} PMA, 2 μg ml^{-1} goat anti-mouse IgG or neither as indicated above and centrifuged at 10 × g for 5 min at 24 °C (or 4 °C for *c*&*d*). In temperature effect experiments similar results were obtained if cells were allowed to settle at 1 × g for 60 min at 4 °C (not shown). Plates were then incubated for 6 min at 37 °C (or 4 °C in *c*&*d*). Unbound cells were removed by washing, and bound cells removed for γ-counting by incubation for 10 min at 37 °C in medium with 10 mM EDTA. Visual inspection of the number of bound cells per well gave the same results as γ-counting. In *a–d*, unbound cells were removed by four complete aspirations of media at assay temperature through an 18 ga. needle as described[49]. In *e* and *f*, unbound cells were removed by flicking media from the plates 8 times with 100 μl added between each wash. Flicking was more effective for thoroughly removing unbound resting T cells which were more difficult to remove due to their small size and it was also quicker, allowing more careful kinetic analysis. Both the aspiration and flicking wash protocols result in higher shear forces than techniques used in earlier studies with cells adhering to ICAM-1 in liposomes or planar membranes[2,49,50]. These differences result in lower percent binding of unstimulated JY cells observed here. Mab used in this study were TS2/4 (native LFA-1 α, IgG1)[51], TS1/18 (native LFA-1 β, IgG1)[51], TS1/22 (LFA-1 α, IgG1)[51], RR1/1 (ICAM-1, IgG1)[21], R6.5 (ICAM-1, IgG2a)[52], CL203 (ICAM-1, IgG1)[53], TS2/9 (LFA-3, IgG1)[51], OKT3 (CD3, IgG2a[54] Ortho Pharmaceuticals, Raritan, New Jersey) and Leu4 (CD3, IgG1)[55] (a gift from R. Evans), 235 (ref. 29) and 38-1 (ref. 25) (CD3 IgMs) were gifts from Drs S. M. Fu and J. Ledbetter, respectively. Anti-CD6 IgM[56] and anti-CD6 IgG2a (JOR-T1, Dr Amador, Stockholm) were obtained from the T-cell panel of the Fourth International Leukocyte Workshop.

to understanding regulation of this adhesion mechanism as it suggests a requirement for membrane fluidity or metabolic energy. As previously demonstrated, binding of JY cells to ICAM-1 was abolished at 4 °C (ref. 2); absence of adhesion was also observed when cells were pretreated with PMA for 30 min at 37 °C before testing adhesion at 4 °C (Fig. 1*c*). Dose-dependent adhesion of JY cells to purified LFA-1 was observed at both 37 °C and 4 °C, although binding at 4 °C was less efficient (Fig. 1*d*). The temperature effect on adhesion to purified ICAM-1 and LFA-1 also shows sidedness, therefore, with binding through cell-surface LFA-1 showing greater sensitivity than adhesion through cell-surface ICAMs.

Definitive evidence for LFA-1/ICAM-1 interaction was obtained by showing that purified LFA-1 protein micelles bind to ICAM-1 on plastic (Fig. 2). Specificity was demonstrated by lack of binding to purified LFA-3, inhibition with EDTA, and inhibition with anti-LFA-1 (TS1/22) and anti-ICAM-1 (RR1/1) monoclonal antibodies, which block cell-cell adhesion. Other antibodies that bind equally well to LFA-1 (TS2/4) or ICAM-1 (CL203) but do not block cell-cell adhesion had no effect. Also, identical results were obtained at 4 °C and 37 °C and thus binding of the purified molecules to each other is not temperature-dependent. Temperature therefore seems to be important for cellular processes that affect the avidity of LFA-1, but not for the function of cell-surface ICAMs, purified LFA-1 or purified ICAM-1.

TCR regulation of LFA-1 avidity

To determine whether regulation of the avidity of LFA-1 was physiologically relevant to antigen-specific T lymphocyte cell-cell interactions, we tested whether it was stimulated by TCR ligation. TCR stimulation triggers phosphatidylinositol turnover and elevates cytoplasmic Ca^{2+} (ref. 23). We studied resting T lymphocytes, rather than cloned T-cell lines, as the latter are already activated by antigen and by maintenance in culture and may be difficult to obtain as a homogeneous resting population. Because resting peripheral blood T cells vary in TCR specificity, bearing different $\alpha\beta$ or $\gamma\delta$ TCR subunits quantitatively associated with the invariant CD3 subunits, an effective way of stimulating them is with a monoclonal antibody against the CD3 component of the TCR complex. Resting T cells were incubated with anti-CD3, washed, and then incubated with a cross-linking anti-IgG, which is required to stimulate hydrolysis of phosphatidyl inositol and Ca^{2+} mobilization[24,25]. Anti-CD3 IgG alone has no effect on T-cell adhesion. Addition of anti-IgG, however, stimulated a significant increase in T-cell adhesion to ICAM-1 substrates, even more than with PMA (Fig. 1*e*). By contrast, there was no effect on adhesion to LFA-1 substrates (Fig. 1*f*). The CD6 antigen, present on T cells in similar amounts as the TCR, was used as a control for nonspecific effects. Cross-linking with an anti-CD6 IgG and anti-IgG did not result in an increase in adhesion (see Fig. 3*a*). The cross-linking anti-IgG could be intact IgG or Fab'_2 (not shown). The TCR-stimulated adhesion to purified ICAM-1 was completely blocked by monoclonal antibodies against LFA-1 α and β subunits and ICAM-1, which have previously been shown to block cell-cell adhesion (Table 1). Antibody TS2/4, which binds to LFA-1 but shows little inhibition of CTL killing[26], and anti-CD2 antibody, however, were not inhibitory.

A remarkable feature of T-lymphocyte interactions with other cells is the rapid progression from strong adhesion to de-adhesion. If the high-avidity state of LFA-1 were reversible, it could provide a mechanism for the unexplained phenomenon of adhesion and de-adhesion cycles. To address this question, we examined the kinetics of TCR-stimulated changes in the LFA-1 avidity state. Remarkably, the high-avidity state of LFA-1 stimulated by TCR cross-linking with anti-CD3 IgG and anti-IgG peaked at 10 min, with complete return to the low-avidity state by 30 min (Fig. 3*a*). By contrast, PMA-stimulated adhesion was maximal at 10 min and remained elevated for at least one hour. Purified T cells show a bimodal expression of LFA-1, as previously reported[27,28], with the two populations differing 2.5-fold in the number of LFA-1 binding sites per cell (Fig. 4). There was no significant change in LFA-1 expression on T cells pretreated with anti-CD3 IgG at 5 or 30 min after addition of anti-IgG. Thus, quantitative changes in LFA-1 surface expression are not involved in regulating either the induced binding at 5 min, or the decreased binding seen at 30 min. Also, addition of PMA to resting T cells that were stimulated by TCR cross-linking for 30 min restored high levels of adhesion within

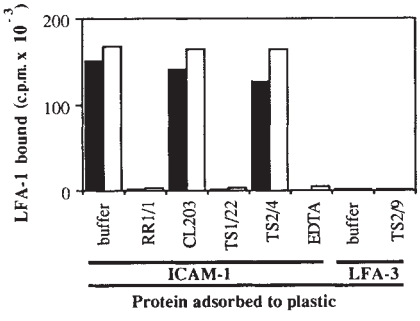


FIG. 2 Binding of LFA-1 protein micelles to ICAM-1 coated plastic. Binding was performed in the presence of the indicated additions at 4 °C (solid bars) or 37 °C (open bars).

METHODS. LFA-1 micelles were prepared by concentrating 2 ml fractions in 1% OG to 50 μl, resuspending in PBS and repeating the concentration step twice to remove residual detergent[57]. Protein concentration could then be determined by Coomassie blue dye binding[57]. LFA-1 micelles were iodinated with lodogen (Pierce, Rockford, Illinois) to a specific activity of 10 μCi μg^{-1} (ref. 57). ^{125}I-labelled LFA-1 micelles (1 μg ml^{-1}) were incubated in microwells with 50 fmol ICAM-1 per well in 50 μl final volume of Hanks balanced salt solution, 10 mM HEPES, 2 mM $MgCl_2$. LFA-3 (ref. 57) and ICAM-1 were adsorbed to plastic as in Fig. 1. Wells were preincubated with purified antibody at 100 μg ml^{-1} prior to adding LFA-1. After 2 h at 37 °C or 4 °C the wells were washed 6 times with assay media, bound LFA-1 was released with 0.1 M NaOH and subjected to gamma particle counting. Data are representative of two experiments. Monoclonal antibodies, as described in Fig. 1.

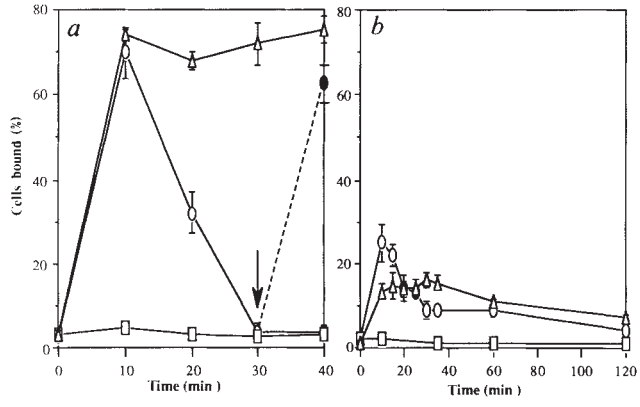


FIG. 3 Kinetics of TCR-stimulated increase in LFA-1 avidity. *a*, Resting T cells were treated with 50 ng ml^{-1} PMA (open triangles) or monoclonal antibody pretreated resting T cells were treated with 1 μg ml^{-1} goat anti-mouse IgG for varying time in suspension before transfer into wells and centrifugation onto an ICAM-1 coated surface. Pretreatment was with 1:500 anti-Leu4/CD3 ascites (open circles) or 1:500 JOR-T1 anti-CD6 ascites (open squares). In one case (arrow) T cells treated with anti-CD3 plus anti-IgG for 30 min at 37 °C were added to wells containing a final concentration of 50 ng ml^{-1} PMA (filled circle). Data is representative of 4 experiments. *b*, as in *a*, except resting T cells were treated with 235 anti-CD3 IgM at 10 μg ml^{-1} (open circles) or 1 μg ml^{-1} (open triangles) or anti-CD6 IgM at 1:500 ascites (open squares). CD6 staining was equivalent to 44% saturation with anti-CD3 monoclonal antibody. These data are representative of two experiments. The indicated time includes the 5 min centrifugation period.

10 min (Fig. 3*a*). This demonstrates that LFA-1 and the adhesion-promoting machinery are still functionally intact after the avidity decrease seen 30 min after TCR cross-linking. In addition, the experimental demonstration of 1.5 cycles of LFA-1 avidity regulation within 40 min indicates that the mechanism is truly capable of cycling, and in a rapid manner. Increased adhesion was also stimulated by anti-CD3 IgM without any need for further cross-linking (Fig. 3*b*), which is consistent with the ability of this anti-CD3 IgM to induce Ca^{2+} flux and phosphatidylinositol turnover in resting T cells[25,29]. The kinetics of the IgM-stimulated adhesion were concentration-dependent. With 10 $\mu g\ ml^{-1}$ anti-CD3 IgM (80% saturation after 30 min at 4 °C, as determined by immunofluorescence flow cytometry), a distinct peak in adhesion to ICAM-1 at 10 min was obtained, which then decreased gradually over 2 h. Absence of an early peak with an even more sustained increase, maximal at 30 min, was obtained with 1 $\mu g\ ml^{-1}$ anti-CD3 IgM (30% saturation after 30 min at 4 °C). No increase in adhesion was seen after addition of anti-CD6 IgM. Stimulation through TCR induces an increase in LFA-1 avidity, therefore, with kinetics dependent upon the degree of cross-linking.

Does the TCR communicate directly with LFA-1, or is it linked to LFA-1 through signalling pathways? Treatment of T cells with the cell-permeable cAMP analogue, dibutyryl cAMP[30-32], before TCR cross-linking or PMA addition, strongly inhibited TCR-stimulated adhesion to ICAM-1 substrates but had no effect on PMA-stimulated adhesion (Fig. 1*e*; Fig. 5). Similar decreases in adhesion triggered by both anti-CD3 IgG plus anti-IgG and anti-CD3 IgM were observed with dibutyryl cAMP (50% of maximal inhibition was achieved at ~1 mM) or the adenylyl cyclase activator forskolin (25 μM) plus the phosphodiesterase inhibitor isobutyl methyl xanthine (IBMX) (0.5 mM) (Fig. 5). These agents were used over ranges previously used to inhibit TCR signalling and were clearly not toxic, as PMA-stimulated adhesion was not decreased (Fig. 1*e*, Fig. 5). Butyrate (5 mM), a control for effects of butyrate released by degradation of dibutyryl cAMP, had no effects on unstimulated, or on cPMA- or TCR-stimulated adhesion to ICAM-1 (data not shown). The protein kinase C inhibitor, staurosporine (5 $\mu g\ ml^{-1}$)[33], completely blocked the PMA-stimulated increase in LFA-1 avidity (Fig. 5), but, surprisingly, it only partially blocked TCR-stimulated adhesion. This illustrates either that protein kinase C can feed into the same pathway as the TCR, but is not a direct intermediate in TCR-stimulated signalling, or that the TCR can stimulate some additional mechanisms. Low concentrations of staurosporine (0.01–0.1 ug ml^{-1}) induced a small (2–5 fold) but significant increase in resting T-cell binding to ICAM-1, indicating that staurosporine may inhibit kinases that have an inhibitory effect on LFA-1 avidity (data not shown). The dramatic and opposite effects of PMA and cAMP, which activate protein kinases C and A, respectively, suggest that LFA-1 avidity is regulated by intracellular signalling pathways. These results support the notion that intracellular second messengers connect the TCR and LFA-1, and the concept that LFA-1 can transduce signals from the cytoplasm to the extracellular environment, an example of 'inside-out signalling'.

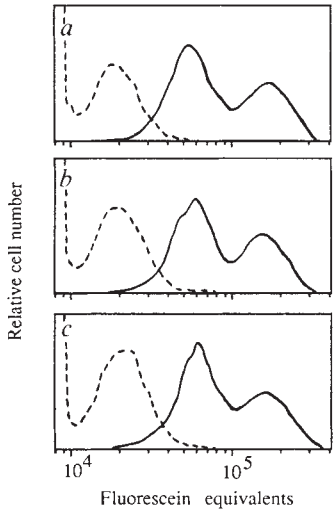


FIG. 4 Expression of LFA-1 on TCR-stimulated resting T cells. Resting T cells were pretreated with OKT3 culture supernatant at 1:10 dilution (~80% saturation) for 30 min at 4 °C and washed. Anti-IgG2a (1 $\mu g\ ml^{-1}$) was either not added (*a*) or added for 5 min (*b*) or 30 min (*c*) at 37 °C before rapid cooling to 0 °C by addition of ice-cold media containing 0.025% NaN_3. Cells were then stained with fluorescein isothiocyanate (FITC) labelled LFA-1 monoclonal antibody (TS1/22) and analysed by immunofluorescence flow cytometry (FITC TS1/22 mAB, solid line; FITC control IgG1, dashed line). Only the relevant portion of the 3 log scale is shown.

TCR regulation of cell–cell adhesion

In interactions between cloned CTL lines and B-lymphoblastoid cell line (BLCL) target cells, both the CD2/LFA-3 and LFA-1/ICAM-1 pathways play a major part in adhesion[16,17]. We examined the effect of TCR cross-linking on the LFA-1/ICAM and CD2/LFA-3 dependent components of conjugate formation between resting T cells and BLCL (Fig. 6*a*, *b*). The BLCL used expressed high levels of ICAM-1 and LFA-3, but were LFA-1^-. We found that unstimulated fresh peripheral blood T cells showed much lower levels of conjugate formation (2–8%) (Fig. 6) than those previously published for cloned CTL (40–80%)[17]. The efficiency of conjugate formation was dramatically increased by TCR cross-linking and the increase in adhesion was completely blocked by anti-LFA-1 monoclonal antibody (Fig. 6*a*). A portion (10–25%) of the basal and stimulated adhesion was blocked by anti-LFA-3 antibody. The TCR-stimulated increase in T-cell adhesion seems to be due mainly to an increase in LFA-1 avidity, with no, or only a minor, contribution from CD2. TCR stimulation of LFA-1 avidity in cell–cell adhesion was transient and was blocked by preincubation of T cells with dibutyryl cAMP (Fig. 6*b*). Thus, the transient increase in LFA-1 avidity stimulated by the TCR is manifested both in binding to purified ICAM-1 and to ICAM-1^+ cells.

Role of LFA-1 avidity in immune response

Based on these findings, we propose the following model for cooperation between TCR and adhesion molecules to mediate antigen-specific recognition. LFA-1 on unactivated cells, such as resting T lymphocytes, is in a low avidity state which may be equivalent to the inactive state of LFA-1 on cells depleted of ATP with sodium azide and 2-deoxyglucose[1,15]. There may be some interaction between LFA-1 in the low avidity state and ICAM-1, but not enough to stabilize cell adhesion. In the absence of antigen, then, the equilibrium governing adherence of T lymphocytes to other cells favours free, mobile T lymphocytes, leading to efficient immune surveillance. On contact with cells bearing specific antigen, TCR ligation generates intracellular signals which lead to energy-dependent conversion of LFA-1 to a high-avidity state and LFA-1/ICAM-dependent adhesion is favoured. Antigen specificity is maintained because the input of energy to convert LFA-1 to the high-avidity stage, whether this energy is used to fuel protein phosphorylation, LFA-1 redistribution or some other mechanism, is controlled by the TCR. Cellular energy expended in converting LFA-1 to a high avidity state drives the adherence–nonadherence equilibrium towards stable adherence, and is analogous to the use of ATP to favour an otherwise energetically unfavourable reaction in intermediary metabolism. LFA-1 is an adhesion servomotor operated by the TCR. As TCR binding to peptide MHC does not have to stabilize cell–cell adhesion but instead triggers an adhesion amplification mechanism, this provides a mechanism for greatly increasing the sensitivity of T cells by lowering the number of TCR–ligand interactions required for antigen

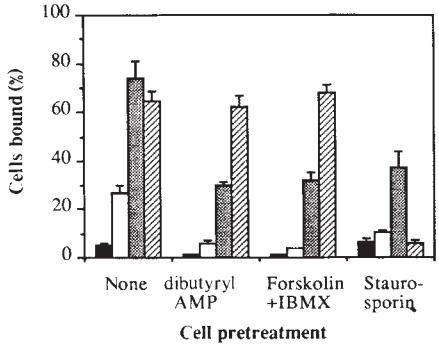


FIG. 5 Effect on LFA-1 avidity increase with pharmacologic agents that act on intracellular cAMP levels or protein kinases. Resting T cells were pretreated with the indicated agents for 15 min at 24 °C before stimulation and assayed for binding to ICAM-1 (1,600 sites μm^{-2}) as in Fig. 1. T cells were unstimulated (solid) or stimulated with anti-CD3 IgM (10 $\mu g\ ml^{-1}$) (open), anti-Leu4/CD3 IgG plus anti-IgG (1:500 Leu4 then 1 $\mu g\ ml^{-1}$ anti-IgG) (shaded), or PMA (50 $ng\ ml^{-1}$) (diagonal stripes). Dibutyryl cAMP (Calbiochem) was dissolved in media and was used at 2 mM. Forskolin (Calbiochem), IBMX (Sigma) and staurosporine (Boehringer) were dissolved in dimethyl sulphoxide (DMSO) and used at 25 μM (0.15% DMSO), 0.5 mM (0.25% DMSO) and 5 $\mu g\ ml^{-1}$ (0.5% DMSO), respectively. DMSO at 0.6% had no effect on adhesion to ICAM-1 (data not shown). Data are representative of two experiments directly comparing these conditions and preliminary experiments to define optimal drug concentrations.

recognition. This view of adhesion strengthening is consistent with the recent observation in murine T-cell clones that LFA-1 and talin, a cytoskeletal protein which co-localizes with a number of integrins at sites of adhesion, redistribute to sites of interaction with antigen-bearing B cells, but not antigen-negative B cells[34]. It is intriguing that redistribution of LFA-1 and talin has been shown to be highly sensitive to low antigen concentrations, as would be expected of adhesion amplification.

The transience of the TCR-stimulated increase in LFA-1 avidity provides a mechanism for regulating the adhesion-de-adhesion cycle. We propose that the TCR triggers a cascade of protein phosphorylation or second messenger production such that an early step in the cascade leads to an increase in LFA-1 avidity and a later step lowers it. The kinetics of the LFA-1 avidity increase triggered by anti-CD3 IgG plus anti-IgG are in good agreement with the kinetics of CTL interaction with target cells. Highly active CTL bind to targets rapidly (0.2–2 min) and can deliver the lethal hit and disengage from the target within an additional 6 min[11,19], but this may be slower for some CTL-target combinations and immunization conditions[19]. It is of interest, therefore, that we have found that the kinetics of avidity regulation are influenced by the number of TCR engaged and the degree of cross-linking. Antigen density may, consequently, influence both the strength and kinetics of adhesion, presumably by affecting the kinetics of the signalling cascade. Duration of adhesion may also be influenced by the level of ICAM expression and whether ICAM-1 or ICAM-2 is the ligand. It is important to remember that as ICAM-1 is inducible by cytokines[1,35], T-cell stimulation could lead to induction of ICAM-1 on antigen-presenting cells, and secondarily alter the kinetics of T-cell interactions.

Although the mechanism of the regulation of LFA-1 avidity is unclear, a change in the conformation of the ICAM binding site or redistribution in the membrane seem most likely; either is compatible with the ability of a non-blocking anti-LFA-1 monoclonal antibody to stimulate LFA-1-dependent homotypic adhesion[36]. Another integrin, gpIIbIIIa, undergoes a well-documented increase in affinity. On unactivated platelets, gpIIbIIIa does not bind fibrinogen but upon activation binds soluble fibrinogen with a dissociation constant, K_d, of 29–45 μM (ref. 37), and no transience is reported. Neutrophil aggregation and binding to endothelium stimulated by chemoattractants is transient[38,39]; there is some evidence that the leukocyte integrin Mac-1 contributes to this[38,40], but estimates vary[40,41]. These phenomena are complicated by chemoattractant-stimulated increases in Mac-1 surface expression, the ability of multiple neutrophil receptors to bind both to endothelial cells and to fragments of C3, rapid shedding from the surface on neutrophil activation of another putative endothelium receptor, Mel-14 (ref. 42), and the generality of the effects on other neutrophil surface receptors that are not integrins[43]. Avidity increases for Mac-1, but not gpIIbIIIa, have been correlated with clustering of receptors in the plane of the membrane[44,45]; surprisingly, redistribution of Mac-1 and increased binding to iC3b are stimulated by phorbol esters but not by a chemoattractant[44], whereas neutrophil aggregation and binding to endothelium are stimulated by both[38,40]. The transience of the increase in avidity of neutrophil Mac-1/CR3 and of the non-integrin CR1 (refs 40, 43), by contrast to the stability of the increase in avidity of LFA-1, stimulated by PMA, suggests that separate mechanisms are involved.

We have defined a mechanism by which lymphocyte adhesion to other cells can be dynamically regulated. The type of inside-out signalling described here in which signals from the cytosol are transduced across the membrane to generate changes in extracellular functions such as adhesion could be of general importance in cell and developmental biology. In the model system studied here, TCR on the entire surface of the cell is ligated, whereas in physiological interactions TCR would be engaged only in the area of cell-cell contact and it would be possible that avidity enhancement would apply only to LFA-1 molecules in or recruited to this area of the cell. Integrins are thought to be important in cell migration as well as adhesion[7]. Spatial gradients of adhesion-molecule avidity on the cell surface could be generated by the same mechanisms as the temporal

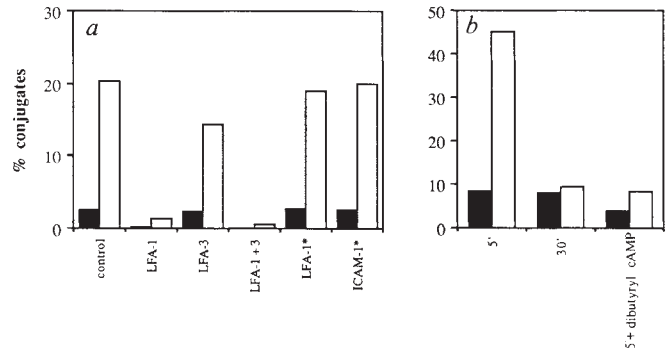


FIG. 6 Regulation of cell-cell adhesion by the TCR. Resting T cells and the genetically LFA-1$^-$ BLCL BBN from patient 1 (ref. 5) were labelled with sulphofluorescein diacetate and hydroethidine, respectively, and conjugates were enumerated by fluorescence flow cytometry[17]. *a*, The resting T cells and BLCL were allowed to co-sediment for 1 h at 4 °C without (solid bars) or with (open bars) OKT3 anti-CD3 monoclonal antibody pretreatment and anti-mouse IgG2a and in the presence of the indicated antibody at 50 $\mu g\ ml^{-1}$. Binding, non-function blocking antibodies are asterisked. LFA-1, LFA-3, LFA-1* and ICAM-1* were TS1/22, TS2/9, TS2/4 and CL203, respectively, and are IgG1, so only OKT3 was crosslinked, as described in Table 1. The pellet was incubated at 37 °C for 5 min and analysed. *b*, Resting T cells were pretreated with OKT3 anti-monoclonal antibody and incubated without or with anti-IgG for 5 or 30 min as indicated (5′ or 30′) and then combined with BLCL and centrifuged at 20g for 5 min at 24 °C, incubated for 5 min at 37 °C and analysed. T cells were pretreated without or with 1 mM dibutyryl cAMP for 15 min at 24 °C before incubation with anti-IgG. METHODS. Cell purification was carried out as in Fig. 1. Labelling of cells was done as described[17]. Resting T cells were combined with BLCL at a 1:2 ratio (5×10^6 cells ml^{-1} total) in 50 μl of assay media. Before analysis the cell pellets were dispersed by adding 0.4 ml balanced salt solution with 1 mM $MgCl_2$ and vortexing for 5 s. Similar results were obtained with a more gentle resuspension.

gradient described here. In models of active cell translocation, it is generally appreciated that a mechanism for de-adhesion is required at the trailing edge of the cell[46,47]. A gradient of integrin avidity from high at the leading edge of a cell to low at the trailing edge could provide a mechanism for de-adhesion at the trailing edge and differential adhesiveness at the leading and trailing edges could drive cell migration. This would be analogous to haptotaxis, the ability of gradients of substrate adhesiveness to promote directed migration of cells[48]. □

Received 17 May; accepted 8 September 1989.

1. Springer, T. A., Dustin, M. L., Kishimoto, T. K. & Marlin, S. D. *A. Rev. Immun.* **5,** 223-252 (1987).
2. Marlin, S. D. & Springer, T. A. *Cell* **51,** 813-819 (1987).
3. Makgoba, M. W. *et al. Nature* **331,** 86-88 (1988).
4. Staunton, D. E., Dustin, M. L. & Springer, T. A. *Nature* **339,** 61-64 (1989).
5. Selvaraj, P. *et al. Nature* **326,** 400-403 (1987).
6. Kishimoto, T. K., O'Connor, K., Lee, A., Roberts, T. M. & Springer, T. A. *Cell* **48,** 681-690 (1987).
7. Hynes, R. O. *Cell* **48,** 549-554 (1987).
8. Kishimoto, T. K. *et al. Adv. Immun.* **46,** 149-182 (1989).
9. Williams, A. F. & Barclay, A. N. *A. Rev. Immun.* **6,** 381-405 (1988).
10. Berke, G. & Levey, R. H. *J. exp. Med.* **135,** 972-984 (1972).
11. Martz, E. *Contemp. Topics Immunobiol.* **7,** 301-361 (1977).
12. Springer, T. A. *et al. Immun. Rev.* **68,** 111-135 (1982).
13. Golstein, P. *et al. Immun. Rev.* **68,** 5-42 (1982).
14. Martz, E., Heagy, W. & Gromkowski, S. H. *Immun. Rev.* **72,** 73-96 (1983).
15. Rothlein, R. & Springer, T. A. *J. exp. Med.* **163,** 1132-1149 (1986).
16. Spits, H. *et al. Science* **232,** 403-405 (1986).
17. Shaw, S. *et al. Nature* **323,** 262-264 (1986).
18. Bierer, B. E. & Burakoff, S. J. *Fedn Proc.* **2,** 2584-2590 (1988).
19. Poenie, M., Tsien, R. Y. & Schmitt-Verhulst, A. *EMBO J.* **6,** 2223-2232 (1987).
20. Patarroyo, M., Beatty, P. G., Fabre, J. W. & Gahmberg, C. G. *Scand. J. Immun.* **22,** 171-182 (1985).
21. Rothlein, R., Dustin, M. L., Marlin, S. D. & Springer, T. A. *J. Immun.* **137,** 1270-1274 (1986).
22. Berrebi, G., Takayama, H. & Sitkovsky, M. V. *Proc. natn. Acad. Sci. U.S.A.* **84,** 1364-1368 (1987).
23. Imboden, J. B. & Stobo, J. D. *J. exp. Med.* **161,** 446-456 (1985).
24. Lerner, A., Jacobson, B. & Miller, R. A. *J. Immun.* **140,** 936-940 (1988).
25. Ledbetter, J. A., June, C. H., Grosmaire, L. S. & Rabinovitch, P. S. *Proc. natn. Acad. Sci. U.S.A.* **84,** 1384-1388 (1987).
26. Ware, C. F. *et al. J. Immun.* **131,** 1182-1187 (1983).
27. Kürzinger, K. *et al. J. Immun.* **127,** 596-602 (1981).
28. Sanders, M. E. *et al. J. Immun.* **140,** 1401-1407 (1988).
29. Bjordahl, J. M., Sung, S. J. Hansen, J. A. & Fu, S. M. *Eur. J. Immun.* **19,** 881-887 (1989).
30. Kaibuchi, K. *et al. Biochem. biophys. Res. Commun.* **104,** 105-112 (1982).
31. Patel, M. D. Samelson, L. E. & Klausner, R. D. *J. biol. Chem.* **262,** 5831-5838 (1987).
32. Klausner, R. D. *et al. J. biol. Chem.* **262,** 12654-12659 (1987).
33. Tamaoki, T. *et al. Biochem. biophys. Res. Commun.* **135,** 397-402 (1986).
34. Kupfer, A. & Singer, S. J. *A. Rev. Immun.* **7,** 309-337 (1989).
35. Dustin, M. L., Staunton, D. E. & Springer, T. A. *Immun. Today* **9,** 213-215 (1988).
36. Keizer, G. D., Visser, W., Vliem, M. & Figdor, C. G. *J. Immun.* **140,** 1393-1400 (1988).
37. Di Minno, G. *et al. Blood* **61,** 140-148 (1983).
38. Buyon, J. P. *et al. J. Immun.* **140,** 3156-3160 (1988).
39. Tonnesen, M. G., Smedly, L. A. & Henson, P. M. *J. clin. Invest.* **74,** 1581-1592 (1984).
40. Lo, S. K., Detmers, P. A., Levin, S. M. & Wright, S. D. *J. exp. Med.* **169,** 1779-1793 (1989).
41. Tonnesen, M. G. *et al. J. clin. Invest.* **83,** 637-646 (1989).
42. Kishimoto, T. K., Jutila, M. A., Berg, E. L. & Butcher, E. C. *Science* **245,** 1238-1241 (1989).
43. Wright, S. D. & Meyer, B. C. *J. Immun.* **136,** 1759-1764 (1986).
44. Detmers, P. A., Wright, S. D., Olsen, E., Kimball, B. & Cohn, Z. A. *J. Cell Biol.* **105,** 1137-1145 (1987).
45. Isenberg, W. M., McEver, R. P., Phillips, D. R., Shuman, M. A. & Bainton, D. F. *J. Cell Biol.* **104,** 1655-1663 (1987).
46. Abercrombie, M. *Expl Cell Res.* **8,** (Suppl.) 188-198 (1961).
47. Weiss, L. *Exp. Cell Res.* **25,** 504-517 (1961).
48. Carter, S. B. *Nature* **213,** 256-260 (1967).
49. Dustin, M. L. & Springer, T. A. *J. Cell Biol.* **107,** 321-331 (1988).
50. Dustin, M. L., Singer, K. H., Tuck, D. T. & Springer, T. A. *J. exp. Med.* **167,** 1323-1340 (1988).
51. Sanchez-Madrid, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **79,** 7489-7493 (1982).
52. Smith, C. W. *et al. J. clin. Invest.* **82,** 1746-1756 (1988).
53. Maio, M. *et al. J. Immun.* **143,** 181-188 (1989).
54. Kung, P. C., Goldstein, G., Reinherz, E. L. & Schlossman, S. F. *Science* **206,** 347-349 (1979).
55. Friedrich, W. *et al. Blood* **59,** 696-701 (1982).
56. Meuer, S. C. *et al. Science* **218,** 471-473 (1982).
57. Dustin, M. L., Olive, D. & Springer, T. A. *J. exp. Med.* **169,** 503-517 (1989).

ACKNOWLEDGEMENTS. We thank Dr S. Marlin for sharing the technique for adsorbing ICAM-1 to plastic, E. Luther for flow cytometry analysis of conjugates, Dr S. Ferrone for monoclonal antibodies and Helen Hourihan and Deborah Wilkinson for manuscript preparation. This work was supported by a grant from the NIH.

The Oct-1 homoeodomain directs formation of a multiprotein–DNA complex with the HSV transactivator VP16

Seth Stern, Masafumi Tanaka & Winship Herr

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

The herpes simplex virus transactivator VP16 participates in the formation of a multiprotein–DNA complex with the ubiquitous octamer-motif-binding factor Oct-1. Complex formation is dependent on specific amino acids in the Oct-1 homoeodomain which are in positions analogous to positive control mutations in helix 2 of the λ phage repressor helix-turn-helix motif, indicating that this structure is an ancient target for protein–protein interactions mediating transcriptional control.

PROTEIN–PROTEIN interactions are fundamental to transcriptional regulation in eukaryotic cells. Such interactions are involved both in formation of a pre-initiation complex between RNA polymerase II and general transcription factors[1,2], and in contacts between sequence-specific DNA-binding transcription factors and the promoter-proximal complex (reviewed in ref. 3). Nevertheless, despite the purification of many transcription factors, specific protein–protein interactions remain largely uncharacterized.

In prokaryotes, however, at least one important interaction between a sequence-specific DNA-binding transcription factor, the phage λ repressor, and RNA polymerase has been well characterized. The λ repressor can positively activate transcription from the promoter P_{RM}, which is the promoter of the gene for λ repressor itself[4]. Amino-acid substitutions in λ repressor that abolish positive regulation of transcription from P_{RM}, but which do not affect either operator binding or negative regulation, have been termed positive-control mutations or *pc* mutations. The *pc* mutations map to a solvent-exposed patch on the surface of the DNA-binding domain of the protein which probably directly contacts the RNA polymerase bound to the promoter[5,6].

In eukaryotes, viral transactivator proteins that do not bind to DNA directly but alter patterns of gene expression, provide model systems to study protein–protein interactions. One of these, the herpes simplex virus (HSV) transactivator VP16, also called Vmw65, VF65 and α-TIF, can form a multiprotein–DNA complex with one or more cellular factors, which include the ubiquitous octamer-motif-binding protein Oct-1 (also called

OTF-1, OBP100 and NFIII)[7-10]. Here we show that formation of this multiprotein-DNA complex is dependent on structural features of Oct-1 which are analogous to features of λ repressor that are involved in the interaction of the repressor with RNA polymerase.

VP16 and Oct-1

The HSV VP16 protein is a late gene product that is incorporated into the virion tegument and, on infection, activates expression of the HSV immediate early (IE) genes[11,12] by binding with Oct-1 to the DNA target sequence TAATGARAT (where R is an unspecified purine)[7-10,13,14]. The TAATGARAT target sequence is a degenerate octamer motif that can also bind to Oct-1 in the absence of VP16 (refs 15 and 16). Both Oct-1 and the closely related lymphoid-specific Oct-2 protein are POU homoeodomain-containing proteins related to the pituitary-specific transcription factor Pit-1/GHF-1 and the nematode *Caenorhabolitis elegans unc-86* gene product[17]. The POU proteins are characterized by an extended region of homology called the POU domain, which is responsible for their ability to bind to DNA[18] and includes a C-terminal homoeodomain and an N-terminal POU-specific region of 75 amino acids[17].

Unlike Oct-2, Oct-1 does not apparently activate transcription of most promoters that contain TATA boxes, as such promoters with synthetic octamer motifs are not active in HeLa cells in which Oct-1 is expressed (see refs 19 and 20, and refs therein). Expression of VP16, however, which itself contains an acidic activation domain[21] analogous to those of the yeast activators GAL4 and GCN4 (refs 22-24), activates the octamer motif in these otherwise inactive promoters in non-lymphoid cells[9,19]. These results indicate that VP16, by virtue of its acidic domain and its ability to form a complex with Oct-1, can reprogramme the transcriptional activity of Oct-1.

Here we show that the Oct-1 homoeodomain directs formation of the complex with VP16. Oct-2 fails to direct the formation of a similar complex[10] because the amino-acid sequence of the Oct-2 homoeodomain differs from that of the Oct-1 homoeodomain at seven out of sixty possible positions. Sequence comparisons[25-28] and a structural analysis[29] indicate that homoeodomains contain a tri-α-helical structure similar to the DNA-binding domains of λ- and 434 repressors[30,31]. When the Oct-1 and Oct-2 homoeodomains are modelled as a DNA-binding domain of λ repressor, the positions of most, if not all, of the seven residues which differ between Oct-1 and Oct-2 lie on the surface of the homoeodomain that faces away from the DNA. Changing the amino-acid residues in Oct-1 at three of these seven positions with those amino acids at the corresponding positions in Oct-2 is sufficient to disrupt the formation of the complex with VP16 without obviously affecting the DNA-binding activity of the protein. Two of these three positions are analogous to the positions of *pc* mutations in the helix-turn-helix motif of λ repressor.

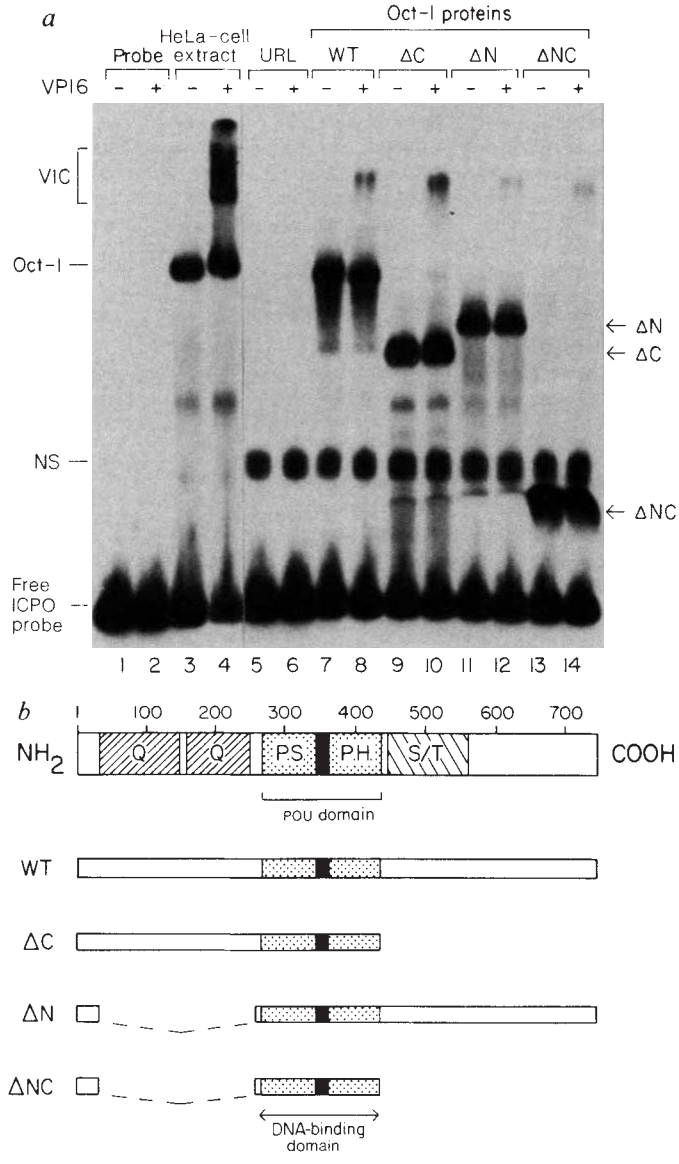


FIG. 1 Formation of a VP16-induced complex (VIC) by Oct-1 deletion mutants. *a*, Gel-retardation assay for VP16 complex formation by HeLa-cell extract or *in vitro* translated Oct-1 protein products and HSV virion lysate with the ICPO probe. Seven different sets of reactions were performed in duplicate, either without (lanes 1, 3, 5, 7, 9, 11 and 13) or with (lanes 2, 4, 6, 8, 10, 12 and 14) HSV virion lysate. Probe alone (lanes 1, 2); HeLa-cell extract (lanes 3, 4); unprogrammed reticulocyte lysate (URL; (lanes 5, 6); wild-type (WT) Oct-1 (lanes 7, 8); (C-terminal deleted Oct-1 (ΔC) (lanes 9, 10); N-terminal deleted Oct-1 (ΔN) (lanes 11, 12); (C- and N-terminal deleted Oct-1 (ΔNC) (lanes 13, 14). The positions of the free ICPO probe, a non-specific complex (NS), full-length Oct-1 (Oct-1), and (VICs) are shown on the left. The positions of the Oct-1-derived ΔC, ΔN and ΔNC complexes are indicated on the right. *b*, Schematic illustration of the Oct-1 proteins produced by *in vitro* translation of RNAs derived from pBSoct-1$^+$ and pBSoct-1$^+$ΔBH. The 743-amino-acid reading frame is boxed, and the glutamine-rich (Q), POU domain (PS, POU-specific region; PH, POU homoeodomain) and serine/threonine-rich (S/T) regions are indicated. Bottom, Truncations and internal deletions used to map the region responsible for VIC formation.

METHODS. The plasmids pBSoct-1$^+$ and pBSoct-1$^+$ΔBH were described previously[32]. Oct-1 constructs were linearized with *Hin*dIII or, in the case of C-terminal deletions, with *Pf*MI. Linearized DNAs (~3.5μg) were used to generate RNA transcripts with T7 RNA polymerase (Stratagene) according to the manufacturer's instructions. Nucleic acids were purified by phenol extraction and ethanol precipitation, and resuspended in 10 μl water. RNAs were translated in a rabbit reticulocyte lysate (Promega), according to the manufacturer's instructions with 1 or 2 μl RNA (~1 μg) per 25 μl reaction volume. The efficiency and quality of *in vitro* translation were monitored by SDS-PAGE of the ^{35}S-labelled products and equivalent amounts of Oct protein were used in each experiment. The buffer conditions for complex formation were as described[10]. Either 5 μg of HeLa cell nuclear extract or 1 μl programmed reticulocyte lysate were incubated alone or in the presence of ~0.06 μg HSV virion lysate containing VP16 (ref. 8) in a 12.5 μl reaction for 30 min at 30 °C. The probe, containing a TAATGARAT homology from the promoter of the *ICPO* gene of HSV, was prepared from the complementary oligonucleotides GATCCCGTGCATGCTAATGATATTCTTTGGG and GATCCCCAAAGAATATCATTAGCATGCACGG, which were treated with kinase, annealed, and cloned into the *Bam*HI site of plasmid pUC119. The probe was excised as an *Eco*RI-*Hin*dIII fragment, 3′ end-labelled using Klenow fragment and 800 Ci mmol^{-1} dNTPs, and gel-purified. A gel slice containing the probe was eluted into 200 μl TE (10 mM Tris-HCl, 1 mM EDTA, pH 8) and 1 μl (10-20,000 c.p.m.) was used in each reaction. After incubation, reactions were loaded directly onto 4%, 40:1 acrylamide:bisacrylamide gels in TBE (25 mM Tris-borate, 0.5 mM EDTA, pH 8.3) and electrophoresed at 12 V cm^{-1} for ~2 h. After electrophoresis the gels were fixed to remove unincorporated [^{35}S]methionine, dried and autoradiographed with two sheets of film to block the ^{35}S decay from the *in vitro* translated products.

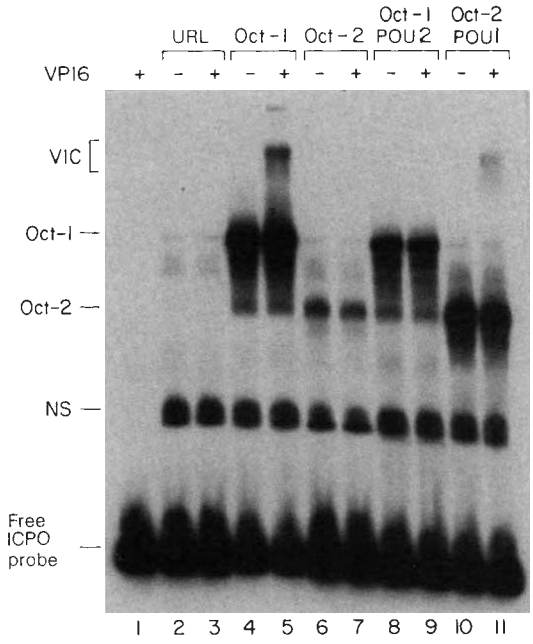


FIG. 2 VP16-induced complex formation by Oct proteins with exchanged POU domains. A gel retardation assay for VP16-induced complex formation by wild-type Oct-1 (lanes 4, 5), Oct-2 (lanes 6, 7), chimaeric Oct-1/POU2 (lanes 8, 9) and Oct-2/POU1 (lanes 10, 11) proteins is shown. Each protein was assayed without (lanes 2, 4, 6, 8, 10) and with (lanes 1, 3, 5, 7, 9, and 11) HSV virion lysate. Lane 1, HSV virion lysate alone; lanes 2 and 3, unprogrammed reticulocyte lysate (URL). The positions of Oct-1, Oct-2 and VP16-induced complexes are indicated on the left. Note that longer exposures of this gel did not reveal any VP16-induced complexes in lanes 7 and 9.
METHODS. The plasmids described here all derive from pBS M13$^+$ (Stratagene). Oct-1 was produced from pBSoct-1$^+$. The Oct-2, Oct-1/POU2 and Oct-1/POU1 proteins were produced from pBS^{B-X-}oct-2(2)2(3)2, pBS^{B-X-}oct-1(2)2(3)1, and pBS^{B-X-}oct-2(2)1(3)2 constructs, respectively. The numbers in parentheses denote silent *Xho*I (2) and *Sal*I (3) restriction sites in the N-terminal-LEQ and C-terminal-RRQ POU-domain encoding sequences, respectively, and the B-X- superscript indicates that the *Bam*HI and *Xba*I sites in the pBS M13$^+$ polylinker were destroyed (by digestion, end-repair and religation). These constructs all contain unique *Xba*I and *Bam*HI restriction sites immediately upstream of the initiation codon (TCTAGAATG), and immediately downstream of the stop codon (TGATGGATCC), respectively. All restriction sites were introduced by oligonucleotide-directed mutagenesis and will be described in detail elsewhere (M.T. and W.H., unpublished results). The *oct-2* initiation (pass-5.5) and termination (pass-3) codons are as previously described[35]. The *oct-1* initiation codon, the first AUG in the *oct-1* sequence, and stop codon are as described previously[32]. The POU domain sequences were exchanged by exchanging *oct-1* and oct-2 sequences between the *Xho*I and *Sal*I sites. To generate templates for transcription *in vitro* by T7 RNA polymerase, plasmids containing the initiation and termination codons for either *oct-1 or oct-2* parental constructs were linearized with *Hin*dIII and *Bam*HI, respectively.

The Oct-1 POU domain and VP16

To assay Oct-1 and VP16 complex formation we used the gel-retardation conditions and a TAATGARAT DNA element described previously[10]. This element was derived from the promoter of the ICP0 gene of HSV, also referred to as IE110K, IE gene 1 and $\alpha 0$, and contains imperfect overlapping octamer (ATGCAAAT) and TAATGARAT motifs (ATGCTAATGATAT). To form VP16-induced complexes, complementary DNAs corresponding to wild-type and mutant Oct-1 were transcribed and translated *in vitro*, and incubated with VP16 supplied by a HSV virion lysate. To prepare the lysate, HSV virions were collected from the growth medium of infected Vero cells and lysed by incubation with Nonidet P40 (NP40) as described by Preston *et al.*[8].

Figure 1*a* shows that incubation of the ICP0 TAATGARAT element (ICP0 probe; lane 1) with a control HeLa cell nuclear extract resulted in the formation of a single main Oct-1 complex (Oct-1; lane 3). As expected, there was no detectable complex formation when the virion lysate containing VP16 was incubated with the probe alone (lane 2), but further addition of HeLa cell nuclear extract resulted in the formation of both the Oct-1 complex and a series of more slowly migrating VP16-induced complexes (VIC; lane 4). Incubation of the probe with an unprogrammed reticulocyte lysate (lane 5) generated little specific Oct-1 complex, but a principal non-specific complex formed (NS); nevertheless, further addition of VP16 to the unprogrammed lysate had no obvious effect (lane 6). A lysate programmed with messenger RNA for Oct-1 (ref. 32), however, produced a major Oct-1 specific complex (lane 7), with faster-migrating minor complexes which probably resulted from internal initiation during *in vitro* translation. Further addition of the HSV virion lysate generated a slow migrating VP16-induced complex (lane 8). The VP16-induced complex can also be formed with VP16 translated *in vitro* (data not shown). In addition, as described previously[8] with Oct-1 derived from HeLa cells, the mobility of the VP16-induced complex is altered by the addition of the VP16-specific monoclonal antibody LP1 (ref. 33), indicating that VP16 is included in the VP16-induced complex (data not shown).

Comparison of lanes 3 and 4 with lanes 7 and 8 (Fig. 1*a*) shows that whereas the DNA-binding activities of Oct-1 from HeLa cell extract and Oct-1 translated *in vitro* are comparable, the VP16-induced complexes formed by the HeLa cell extract are both more numerous and complex. This difference has not so far been accounted for, but could in part have been attributable to the presence of additional HeLa cell proteins in the complex[10]. If such factors are required for complex formation, they are apparently also present in the rabbit reticulocyte lysate, although probably in limiting amounts–hence the relatively weak VP16-induced complex that is observed. In all the studies presented here, except those of Fig. 4, an Oct-1-depleted HeLa cell extract was not included to enhance formation of the VP16-induced complexes[10]. This is because the background level of

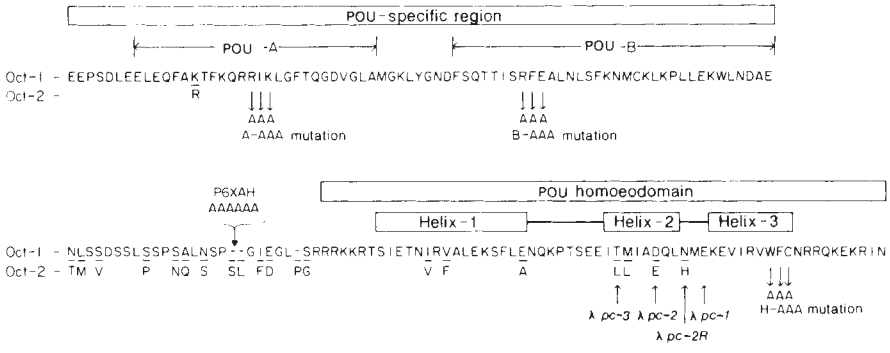


FIG. 3 Sequence comparison of the Oct-1 and Oct-2 POU domains. The locations of the POU-specific A and B regions and the POU homoeodomain are shown above the amino-acid sequence of Oct-1 (ref. 32). Oct-2 residues[35,43,44] that differ from Oct-1 are shown below the Oct-1 sequence, as are the positions of the A-AAA, B-BBB and H-AAA missense mutations, and the P6XAH insertion mutation. The three α-helices predicted to exist in the homoeodomain are indicated, as well as the positions of λ repressor *pc 1-3* mutations and the *pc-2* second site revertant *pc-2R*.

complex formation by unprogrammed reticulocyte lysate also increases, probably because of residual Oct-1 protein in the depleted extract. Regardless of the detailed composition of the complex, however, the results in Fig. 1*a* show that Oct-1 translated *in vitro* has the capacity to participate in a VP16-induced complex, and is therefore probably responsible for the OTF-1 activity described by Fletcher *et al*[34] and Gerster and Roeder[10].

Oct-1 can be divided into three regions: an N-terminal glutamine-rich region, a C-terminal serine/threonine-rich region, and a central POU domain composed of the POU-specific region and homoeodomain (Fig. 1*b*). We used previously characterized Oct-1 deletion mutants[32] to map the domain(s) of Oct-1 involved in complex formation with VP16. *In vitro* translated Oct-1 proteins with either their C-termini or N-termini deleted, or both (leaving a POU domain) were incubated with the ICP0 probe in the absence or presence of HSV virion lysate. All of these proteins bound to the ICP0 probe and produced complexes with mobilities that were approximately proportional to their relative molecular masses (Fig 1*a*; lanes 9, 11 and 13). Furthermore, as with full-length Oct-1, all of the deletion constructs formed a slowly migrating VP16-induced complex (lanes 10, 12 and 14). The mobility of the VP16-induced complexes shifts slightly as the size of the Oct-1 derivative changes, indicating that the different-length Oct-1 proteins are a part of the complex (compare lanes 12 and 14); the small magnitude of the shift, however, is consistent with other cellular proteins being present in the complex. The ability of the deleted Oct-1 proteins to complex with VP16 shows that the Oct-1 POU domain is sufficient to participate in, and direct the formation of, the protein-DNA complex that contains VP16. This result was unexpected, because Oct-2, which does not direct the formation of a complex with VP16 (ref. 10), contains a very similar POU domain.

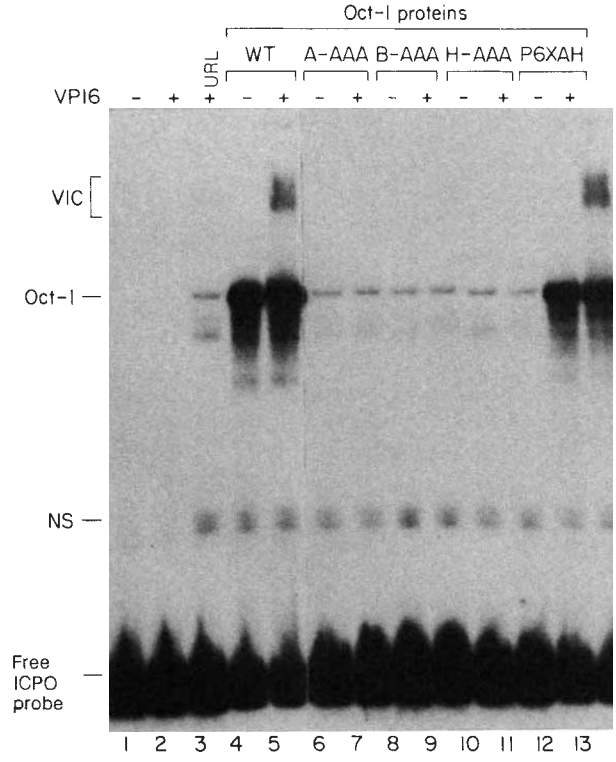


FIG. 4 VP16-induced complex (VIC) formation by Oct-1 POU domain mutant proteins. A gel retardation assay for VIC formation by wild-type Oct-1 (lanes 4, 5), the POU-specific A-AAA (lanes 6, 7), B-AAA (lanes 8, 9) and H-AAA (lanes 10, 11) triple amino-acid substitution mutants, and the Oct-1 insertion mutant with a six alanine insertion in the POU-linker (P6XAH; lanes 12, 13) is shown. Oct-1-depleted HeLa-cell extract was added to every reaction, and each protein was assayed without (lanes 1, 4, 6, 8, 10, and 12) and with (lanes 2, 3, 5, 7, 9, 11, and 13) HSV virion lysate. Lane 1, depleted extract only, lane 2, added VP16; lane 3, added VP16 and unprogrammed reticulocyte lysate (URL). Note that the addition of unprogrammed reticulocyte lysate to the Oct-1-depleted HeLa cell extract results in a small amount of residual Oct-1 binding activity and, in the presence of HSV virion lysate, a VIC is clearly evident after longer exposure of the film. The positions of Oct-1 complexes and VICs are shown on the left.

METHODS. Plasmids pBSoct-1$^+$A-AAA, pBSoct-1$^+$B-AAA, pBSoct-1$^+$H-AAA, and pBSoct-1$^+$P6XAH have been described[18]. HeLa cell nuclear extract was depleted for octamer-binding activity by repeated incubation with biotinylated octamer motif-containing DNA fragements linked to streptavidin-agarose beads. Depleted extract (1 μl) corresponding to 5 μg of proteins before depletion, was added to each reaction.

The inability of Oct-2 to form a complex with VP16 could be intrinsic to its POU domain, or alternatively, the Oct-2 POU domain might be inhibited from forming a complex by the flanking N- and C-terminal regions of Oct-2. To distinguish between these two possibilities we constructed full-length chimaeric Oct-1 and Oct-2 proteins in which the POU domains were exchanged (Oct-1/POU2 and Oct-2/POU1), and assayed their capacity to form a complex with VP16 (Fig. 2). Consistent with the results with purified OTF-2 (ref. 10), incubation of the ICP0 probe with Oct-2 that was produced *in vitro* from the cloned *oct-2* gene[35] resulted in an Oct-2 complex, which migrates faster than the Oct-1 complex (lane 6), and addition of virion lysate did not result in the formation of a VP16-induced complex (lane 7). Under the gel retardation conditions described here, Oct-2 binds less effectively to the TAATGARAT probe than does Oct-1. Nevertheless, the results described below (for example, see Fig. 6) show that the inability of Oct-2 to complex with VP16 is not solely due to this weaker affinity of Oct-2 for the TAATGARAT probe. The chimaeric Oct-1/POU2 protein formed a complex with the probe with a mobility identical to that of the complex formed with the probe and wild-type Oct-1 (lane 8), but like Oct-2, did not form a complex with VP16 (lane 9). The chimaeric Oct-2/POU1 protein, however, produced an Oct-2-type complex with the probe (lane 10), but also formed a complex with VP16 (lane 11). Thus, the ability or inability to form a complex with VP16 is a property intrinsic to the POU domains of Oct-1 and Oct-2, respectively.

The Oct-1 homoeodomain and VP16

Figure 3 compares the sequences of the POU domains of Oct-1 and Oct-2, showing the positions of the POU-specific region and the homoeodomain. A 'linker' of 24 or 27 amino acids separates these two regions in Oct-1 and Oct-2, respectively. The POU-specific region can be subdivided into A and B segments, based largely on the position of three additional amino acids found in the POU domain of the *unc-86* gene product[36]. Oct-1 and Oct-2 differ at 1 out of 75, 12 out of 24, and 7 out of 60 positions in the POU-specific, linker and homoeodomain regions, respectively (Fig. 3).

We have previously used triple alanine substitutions in the POU-specific A region (A-AAA mutation) and B region (B-AAA mutation), and in the POU homoeodomain (H-AAA mutation), to show that these three regions are all necessary for effective DNA binding *in vitro*[18]. The poorly conserved linker region, however, is apparently not intimately involved in DNA binding because the six alanine insertion in this region (P6XAH mutation) (see Fig. 3) does not affect DNA-binding activity[18]. Here we used these four POU-domain mutations to investigate whether VP16 can functionally replace one of the mutated regions, or whether VP16 recognizes the Oct-1 linker, the POU domain region of greatest sequence divergence between Oct-1 and Oct-2. VP16 does not activate any latent DNA-binding activity in the A-AAA, B-AAA or H-AAA mutants (Fig. 4; lanes 6–11), and VP16-induced complex formation is unaffected by the linker insertion-mutation (compare lanes 5 and 13). These results indicate that VP16 cannot functionally replace the POU-specific region or the homoeodomain for DNA binding, and that the divergent linker region is relatively unimportant for VP16-induced complex formation.

To test the relative contributions of the POU-specific region and the homoeodomain to the formation of the VP16-induced complex, we used Oct-1 and Oct-2 proteins in which either their homoeodomain regions or POU-specific regions had been exchanged. When the Oct-1 POU-specific region and linker region were replaced by the corresponding Oct-2 sequences, there was no obvious effect on complex formation (Fig. 5*a*; compare lanes 2 and 4). This shows that VP16 does not distinguish between the Oct-1 and Oct-2 POU-specific sequences; but, because these two regions differ at only one position, it is nevertheless possible that the POU-specific region makes protein–protein contacts essential to form a VP16-induced complex. To test further the contribution of the POU-specific sequences to VP16-induced complex formation we made an Oct-1 protein containing most of the POU-specific region (the A and B regions) and all of the linker sequences from the other mammalian POU protein, Pit-1/GHF-1 (refs 37 and 38). Pit-1/GHF-1 binds to octamer-related sequences[39], but its POU-specific region and homoeodomain differ from those of Oct-1 at 39% and 43% of the positions, respectively; furthermore, the Pit-1/GHF-1 linker region is only 15 amino acids long and is totally divergent from that of Oct-1. Oct-1 that carried the POU-specific and linker sequences of Pit-1/GHF-1 (Oct-1/PS-P), however, could still direct complex formation with VP16, albeit at a reduced level, which may reflect a reduced affinity for the ICP0 probe (Fig. 5*a*; compare Oct-1 complexes in lanes 1 and 5). Although Oct-1/PS-P could still form a complex, an Oct-1 chimaera that carried nearly the entire Pit-1/GHF-1 POU domain (Oct-1/POU-P; see Fig. 5 legend) was unable to direct effective complex formation (lane 8). These results, together with those obtained with the Oct-1 POU domain mutants, indicate that the POU-specific and linker regions themselves do not contain determinants for the formation of a VP16-induced complex, but are nonetheless required because they are involved in DNA binding.

Figure 5*b* shows that the Oct-1 homoeodomain contains determinants for VP16-induced complex formation and is sufficient to activate the formation of a VP16-induced by Oct-2. When the Oct-2 homoeodomain was placed in Oct-1 (Oct-1/Ho-2), the VP16-induced complex disappeared (compare lanes 4 and 6), but when the Oct-1 homoeodomain was placed in Oct-2 (Oct-2/Ho-1), VP16-induced complex formation was activated (compare lanes 8 and 10), although at a level less than that with wild-type Oct-1 (compare lanes 4 and 10). This indicates that Oct-1 sequences outside the homoeodomain influence VP16-induced complex formation. These results show that the seven amino acids that differ between the Oct-1 and Oct-2 homoeodomains are sufficient to allow VP16 to discriminate between these two proteins.

Helix swap

Figure 3 shows the positions of the three predicted α-helices in the Oct-1 homoeodomain sequence, and Fig. 6*a* shows how they might exist in three dimensions should they be arranged like the λ repressor DNA-binding domain. The amino acids that differ between Oct-1 and Oct-2 are shown in both diagrams.

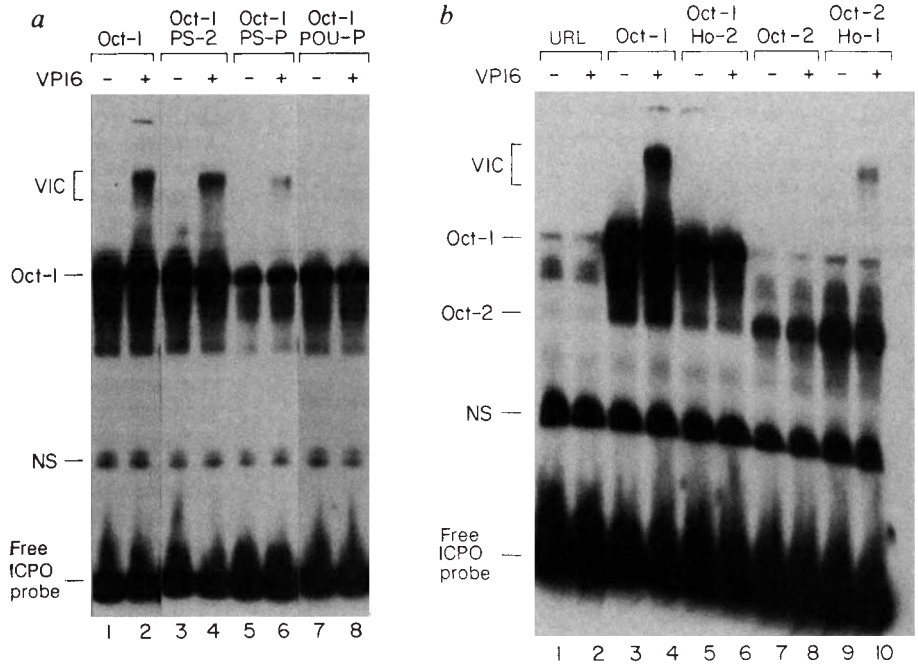


FIG. 5 VIC formation by Oct proteins with exchanged POU-specific regions or homoeodomains. *a*, Gel retardation assay of VIC formation by wild-type Oct-1 (lanes 1, 2), and Oct-1 carrying the Oct-2 POU-specific A and B regions (Oct-1/PS-2; lanes 3, 4), the Pit-1/GHF-1 POU-specific A and B regions (Oct-1/PS-P; lanes 5, 6) or Pit-1/GHF-1 POU domain sequences (Oct-1/POU-P; lanes 7, 8) Each protein was assayed without (lanes 1, 3, 5, 7) and with (lanes 2, 4, 6, 8) HSV virion lysate. *b*, Gel retardation assay for VP16-induced complex formation by wild-type Oct-1 (lanes 3, 4), Oct-1 carrying the Oct-2 homoeodomain (Oct-1/Ho-2; lanes 5, 6), wild-type Oct-2 (lanes 7, 8) and Oct-2 carrying the Oct-1 homoeodomain (Oct-2/Ho-1; lanes 9, 10). Each protein is assayed without (lanes 1, 3, 5, 7, and 9) and with (lanes 2, 4, 6, 8, 10) HSV virion lysate. Lanes 1 and 2 have unprogrammed reticulocyte lysate (URL) The positions of Oct-1 and Oct-2 complexes and VICs are shown on the left. Note that longer exposure of this gel did not reveal any VP16-induced complex in lanes 6 and 8.

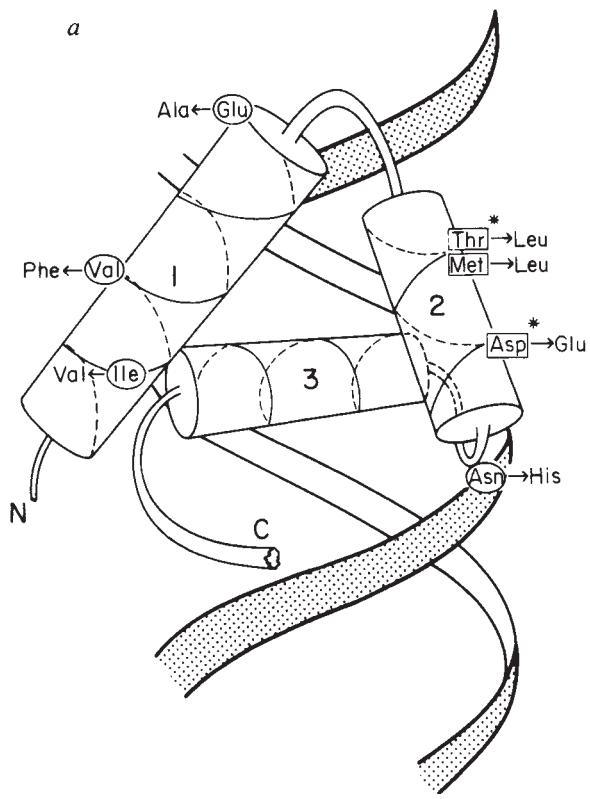
METHODS. The Pit-1/GHF-1 POU domain sequences[38] were cloned as follows. Complementary DNAs were produced from total GH3 pituitary cell RNA with random primers and the POU domain sequence was amplified by the polymerase chain reaction using Pit-1/GHF-1-specific primers that simultaneously created *Xho*I (with an overlapping *Sac*I site—GAGCTCGAG—as in the *oct-1* and *oct-2* genes) and *Sal*I sites at positions equivalent to the *Xho*I (2) and *Sal*I (3) sites in the *oct-1* gene described in Fig. 2 legend. In the pBSoct-1 parental construct used for the constructions described below (called pBS$^{\Delta}$oct-1), the polylinker *Bam*HI site is present and the sequences between the polylinker *Xba*I site and the *Xba*I site flanking the initiation codon have been deleted, but the 3′ termination codon *Bam*HI site remains, as described in Fig. 2 legend. The Pit-1/GHF-1 POU domain sequences between the engineered *Xho*I and *Sal*I sites were transferred to the pBS$^{\Delta}$oct-1(2)1(3)1 construct to create pBS$^{\Delta}$oct-1(2)P(3)1 (Oct-1/POU-P). In this construct the *Pit-1/GHF-1* POU domain sequences 5′ of the POU-specific A region A (coding for 13 amino acids) and 3′ of the *Sal*I site in the homoeobox (coding for 7 amino acids with three differences between Oct-1 and Pit-1/GHF-1) derive from the *oct-1* sequence. To exchange POU-specific regions and POU homoeoboxes a silent *Nar*I restriction site was introduced in the sequences encoding the N-terminal of the homoeodomain (RRR sequence in *oct-1* and *oct-2*, and RKR sequence in *oct-1/POU-P*) in pBS$^{\Delta}$oct-1(2)1(3)1, pBS^{B-X-}oct-2(2)2(3)2 and pBS$^{\Delta}$oct-1(2)P(3)1 by oligonucleotide-directed mutagenesis. The wild-type sequences were mutated to CGGCGCCGG (*oct-1*), CGGCGCCGC (*oct-2*), AGGCGCCGG (*Pit-1/GHF-1*) (the mutations are underlined) creating pBS$^{\Delta}$oct-1(2)1(N)1(3)1, pBS^{B-X-}oct-2(2)2(N)2(3)2, and pBS$^{\Delta}$oct-1(2)P(N)P(3)1, respectively; the mutations are silent in *oct-1* and *oct-2*, whereas the *Pit-1/GHF-1* wild-type RKR sequence is altered to RRR. Oct-1/PS-2 [pBS$^{\Delta}$oct-1(2)2(N)1(3)1], Oct-1/PS-P [pBS$^{\Delta}$oct-1(2)P(N)1(3)1], Oct-1/Ho-2 [pBS$^{\Delta}$oct-1(2)1(N)2(3)1] and Oct-2/Ho-1 [pBS^{B-X-}oct-2(2)2(N)1(3)2] were all constructed by transferring the appropriate *Xho*I/*Nar*I or *Nar*I/*Sal*I sequences between the parental *Nar*I mutant constructs. To generate RNAs for *in vitro* translation, the pBSoct-1 constructs (including *Pit-1/GHF-1* derivatives) and pBSoct-2 constructs were linearized with *Hin*dIII and *Bam*HI, respectively.

Three amino-acid differences are found in helix-1, three in helix-2, and the seventh in the turn between helices-2 and -3 (Fig. 3). The position of three λ repressor *pc* mutations (*pc-1, pc-2* and *pc-3*), which interfere with the interaction of repressor and RNA polymerase but do not affect DNA-binding activity, are also shown (Fig. 3)[5,6], as is the position of a second site revertant of *pc-2, pc-2R* (ref. 5). Many or all of the seven positions that differ between Oct-1 and Oct-2 apparently lie on the surface of the homoeodomain (Fig. 6*a*). Furthermore, *pc-2* and *pc-3* map to positions that differ between Oct-1 and Oct-2 (see asterisks in Fig. 6*a*), indicating a parallel between Oct-1-directed VP16-complex formation and the interaction of λ repressor with RNA polymerase. To test the possibility of such a parallel, we exchanged the helices-2 of Oct-1 and Oct-2 by making three amino-acid substitutions in each protein.

Such a substitution (Oct-1/He_2-2) disrupts the formation of the VP16 complex (Fig. 6*b*; compare lanes 4 and 6); the longer exposure of the gel reveals that the formation of a residual complex between VP16 and the Oct-1/He_2-2 mutant is less than 10% of the complex formed between wild-type Oct-1 and VP16. But the reciprocal construct, in which the helix-2 of Oct-2 was replaced by the Oct-1 sequence (Oct-2/He_2-1), remained inactive (lane 10), showing that these three residues are necessary but not sufficient for complex formation. Although the Oct-1/He_2-2 mutant affects VP16-induced complex formation greatly, its DNA-binding activity seems to be identical to that of wild-type Oct-1 (compare the Oct-1 complexes in lanes 3 and 4 with those in lanes 5 and 6). Thus the interactions of Oct-1 with DNA and VP16 can apparently be dissociated. This result indicates that the reduction in VP16-induced complex formation by this chimaeric protein is attributable to a reduction in its capacity to direct complex formation with VP16, and is not attributable to a reduction in its affinity for the ICP0 probe.

The homoeodomain mediates positive control

Our results indicate that the Oct-1 homoeodomain is analogous to the helix-turn-helix motif of λ repressor in that it is simultaneously responsible for interaction with DNA and with other protein factors implicated in transcriptional control. Specifically, as with λ repressor *pc* mutations, mutations in the putative helix-2 of the Oct-1 homoeodomain can disrupt a protein–protein interaction that is important for transcriptional activation without an observable affect on DNA-binding activity. The



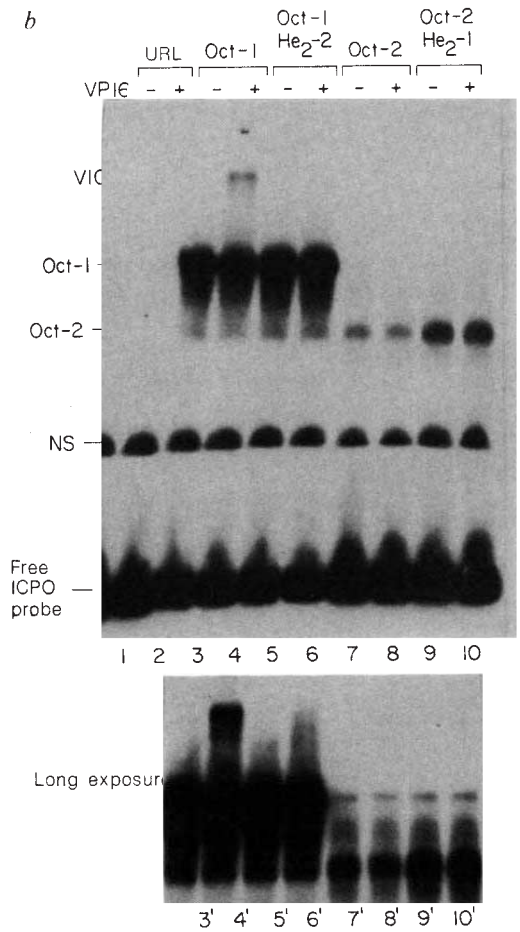


FIG. 6 Homoeodomain model and VP16-induced complex formation by Oct proteins with exchanged helix-2. *a*, Diagram of the tri-α-helical DNA-binding motif of λ repressor[30] and DNA (behind) as a model for the POU homoeodomain (S. Harrison, personal communication). The three α-helices are drawn as cylinders, with the α-carbon backbone indicated by solid (facing viewer) and dashed (hidden from viewer) lines. The seven amino-acid positions that differ between Oct-1 and Oct-2 are circled or boxed, with the Oct-1 residues shown inside and an arrow pointing to the Oct-2 residue. Boxed positions were exchanged in the helix-swap experiment, and asterisks indicate those positions in the homoeodomain that are analogous to λ repressor *pc-2 and pc-3* mutations. *b*, Gel retardation assay for VP16-induced complex formation by wild-type Oct-1 (lanes 3, 4), Oct-1 carrying Oct-2/helix-2 (Oct-1/He_2-2; lanes 5, 6), wild-type Oct-2 (lanes 7, 8) and Oct-2 carrying the Oct-1/helix-2 (Oct-2/He_2-1; lanes 9, 10). Each protein was assayed without (lanes 1, 3, 5, 7, and 9) and with (lanes 2, 4, 6, 8, and 10) HSV virion lysate. The positions of Oct-1 and Oct-2 complexes and VICs are shown on the left. Bottom, Longer exposure of the Oct complexes and VICs.

METHODS. Proteins with exchanged helices-2 were constructed by creating the mutant pBS^{Δ}oct-1(2)1(3)1 and $pBS^{B\text{-}X}$-oct-2(2)2(3)2 constructs, pBS^{Δ}oct-1(2)1$He_2$2(3)1 and $pBS^{B\text{-}X}$-oct-2(2)2$He_2$1(3)2, respectively, each with the three amino-acid changes in helix-2 as indicated in *a* by oligonucleotide-directed mutagenesis. Oligonucleotides containing the sequences CTTCTGATTGCTGAG and ACGATGATCGCCGATCA were used for Oct-1 and Oct-2, respectively, and the mutations are underlined.

inability of Oct-2/He_2-1 mutant to direct VP16 complex formation, however, shows that additional homoeodomain residues are required for activation of complex formation. The analogy to λ repressor indicates that the variable residue in the turn between helices-2 and -3 of Oct-1 and Oct-2 (see Fig. 3) may be critical for the interaction with VP16, because this is the position in λ repressor of the *pc-2* second site revertant, *pc-2R* (see Fig. 3), indicating that amino acids at this position may also be involved in positive control. Conversely, because variable residues in helix-1 of Oct-1 and Oct-2 may also be involved in complex formation, by analogy, similar positions in λ repressor may also be required for positive control.

Although the experiments described here show that the homoeodomain contains determinants that are sufficient to both inactivate and activate VP16-induced complex formation, they do not rule out the possibility that the POU-specific region also makes important protein-protein contacts with components of the VP16 complex. If this is the case, then this interaction probably involves residues in the POU-specific region that were not altered in the Oct-1 chimaera that contained Pit-1/GHF-1 POU-specific sequences (Fig. 5*a*), because this protein was still able to form the complex.

The amino-acid differences between Oct-1 and Oct-2 in helix-2 that affect formation of the VP16-induced complex follow no obvious pattern; there is a conservative aspartic to glutamic acid substitution, and two leucines replace the threonine and methionine residues found in Oct-1 (Fig. 6*a*). This is in contrast to the λ repressor *pc* mutations, which reduce the acidic character of the helix-2 region, an observation that indicated a parallel between positive control by λ repressor and the acidic activation domains of eukaryotic transcription factors[22,23]. VP16-mediated transcriptional activation also seems to involve an acidic activation domain, but in this case VP16 acts as an adaptor that effectively converts the Oct-1 homoeodomain to an acidic activation domain by a protein-protein association that itself no longer depends on acidic residues *per se*.

Implications for transcriptional regulation

Does the Oct-1 homoeodomain direct protein-protein interactions to regulate normal cellular transcription? Perhaps, to ensure a productive HSV infection, VP16 mimics a host-cell factor essential for cell viability, a cellular VP16 analogue. For example, a cellular VP16 analogue could be responsible for the octamer-dependent cell-cycle regulation of the histone H2B promoter[40]. Because it is not known whether VP16 itself, or instead a cellular factor, interacts directly with the Oct-1 homoeodomain, the exact role of a cellular VP16 analogue is unclear. If VP16 does not interact directly with the Oct-1 homoeodomain, then the results described here indicate that the cellular factor involved in complex formation represents a homoeodomain-binding protein. If VP16 does interact directly with the Oct-1 homoeodomain, it may be possible to design Oct-1 mutants that selectively interfere with VP16-induced complex formation without disrupting interactions with other cellular factors, thus serving to prevent HSV infection.

VP16 may interact with the Oct-1 DNA-binding domain because it must ensure transcription of the IE genes on HSV infection. Transcription of IE genes depends on the incorporation of Oct-1, which normally binds preferentially to the cellular octamer motif[15], in a VP16-induced complex that assembles on TAATGARAT sequence elements. VP16 may interact with the DNA-binding domain of Oct-1 to act as an allosteric effector that alters the DNA-binding specificity of Oct-1.

The ability of VP16 to discriminate between two proteins, Oct-1 and Oct-2, that recognize the same *cis*-acting element, shows how non-DNA-binding transcription factors can differentially influence transcription by two proteins that bind to the same DNA sequence. Like Oct-1 and Oct-2, *Drosophila* homoeodomain proteins also bind to similar sequences[41,42]. The results described here therefore indicate that protein-protein contacts that involve homoeodomains themselves could be important for developmental regulation. Indeed, the homoeodomain could have been selected during evolution to regulate development because its structure is amenable to protein-protein interactions. Consistent with this hypothesis, positions in helix-2 of homoeodomains in the products of the Antennapedia class of genes are hypervariable[27], a feature that could facilitate the discrimination among these very similar homoeodomains by a VP16-like factor.

We have shown protein-protein interactions involving the homoeodomain can alter patterns of gene expression. The utilization of the helix-turn-helix motif for DNA-binding functions and *trans*-acting activation functions, in a mammalian transcription factor and in λ repressor, provides a striking example of the conservation of protein structure and function between prokaryotes and man. □

Received 7 June; accepted 14 September 1989.

1. Van Dyke, M. W., Roeder, R. G. & Sawadogo, M. *Science* **241,** 1335-1338 (1988).
2. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. *Cell* **56,** 549-561 (1989).
3. Ptashne, M. *Nature* **335,** 683-689 (1988).
4. Ptashne, M. *et al. Cell* **19,** 1-11 (1980).
5. Hochschild, A., Irwin, N. & Ptashne, M. *Cell* **32,** 319-325 (1983).
6. Hawley, D. K. & McClure, W. R. *Cell* **32,** 327-333 (1983).
7. McKnight, J. L. C., Kristie, T. M. & Roizman, B. *Proc. natn. Acad. Sci U.S.A.* **84,** 7061-7065 (1987).
8. Preston, C. M., Frame, M. C. & Campbell, M. E. M. *Cell* **52,** 425-434 (1988).
9. O'Hare, P. & Goding, C. R. *Cell* **52,** 435-445 (1988).
10. Gerster, T. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **85,** 6347-6351 (1988).
11. Batterson, W. & Roizman, B. *J. Virol.* **46,** 371-377 (1983).
12. Campbell, M. E. M., Palfreyman, J. W. & Preston, C. M. *J. molec. Biol.* **180,** 1-19 (1984).
13. Mackem, S. & Roizman, B. *Proc. natn. Acad. Sci. U.S.A.* **79,** 4917-4921 (1982).
14. Preston, C. M., Cordingley, M. G. & Stow, N. D. *J. Virol.* **50,** 708-716 (1984).
15. Baumruker, T., Sturm, R. & Herr, W. *Genes Dev.* **2,** 1400-1413 (1988).
16. ApRhys, C. M. J., Ciufo, D. M., O'Neill, E. A., Kelly, T. J. & Hayward, G. S. *J. Virol.* **63,** 2798-2812 (1989).
17. Herr, W. *et al. Genes Dev.* **2,** 1513-1516 (1988).
18. Sturm, R. A. & Herr. W. *Nature* **336,** 601-604 (1988).
19. Tanaka, M., Grossniklaus, U., Herr, W. & Hernandez, N. *Genes Dev.* **2,** 1764-1778 (1988).
20. Schaffner, W. *Trends Genet.* **5,** 37-39 (1989).
21. Triezenberg, S. J., Kingsbury, R. C. & McKnight, S. L. *Genes Dev.* **2,** 718-729 (1988).
22. Hope, I. A. & Struhl, K. *Cell* **46,** 885-894 (1986).
23. Ma, J. & Ptashne, M. *Cell* **48,** 847-853 (1987).
24. Hope, I. A., Mahadevan, S. & Struhl, K. *Nature* **333,** 635-640 (1988).
25. Laughon, A. & Scott, M. P. *Nature* **310,** 25-31 (1984).
26. Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., Robertis, E. M. D. & Gehring, W. J. *Nature* **310,** 70-71 (1984).
27. Scott, M. P., Tamkun, J. W. & Hartzell III, G. W. *BBA Reviews on Cancer* (in the press).
28. Garcia-Blanco, M. A., Clerc, R. G. & Sharp, P. A. *Genes Dev.* **3,** 739-745 (1989).
29. Otting, G. *et al. EMBO J.* **7,** 4305-4309 (1988).
30. Jordon, S. R. & Pabo, C. O. *Science* **242,** 893-899 (1988).
31. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. *Science* **242,** 899-907 (1988).
32. Sturm, R. A., Das, G. & Herr, W. *Genes Dev.* **2,** 1582-1599 (1988).
33. McLean, C. *et al. J. gen. Virol.* **63,** 297-305 (1982).
34. Fletcher, C., Heintz, N. & Roeder, R. G. *Cell* **51,** 773-781 (1987).
35. Clerc, R. G., Corcoran, L. M., LeBowitz, J. H., Baltimore, D. & Sharp, P. A. *Genes Dev.* **2,** 1570-1581 (1988).
36. Finney, M., Ruvkun, G. & Horvitz, H. R. *Cell* **55,** 757-769 (1988).
37. Bodner, M. *et al. Cell* **55,** 505-518 (1988).
38. Ingraham, H. A. *et al. Cell* **55,** 519-529 (1988).
39. Nelson, C., Albert, V. R., Elsholtz, H. P., Lu, L. I.-W. & Rosenfeld, M. G. *Science* **239,** 1400-1405 (1988).
40. LaBella, F., Sive, H. L., Roeder, R. G. & Heintz, N. *Genes Dev.* **2,** 32-39 (1988).
41. Desplan, C., Theis, J. O'Farrell, P. H. *Cell* **54,** 1081-1090 (1988).
42. Hoey, T. & Levine, M. *Nature* **332,** 858-861 (1988).
43. Muller, M. M., Ruppert, S., Schaffner, W. & Matthias, P. *Nature* **336,** 544-551 (1988).
44. Scheidereit, C., Cromlish, J. A., Gerster, T., Kawakami, K., Balmaceda, C.-G., Currie, R. A. & Roeder, R. G. *Nature* **336,** 551-557 (1988).

ACKNOWLEDGEMENTS. We thank M. Cleary for help in preparing the HSV virion lysate; T. Baumruker for his involvement in the initial phase of this study; R. Clerc and P. Sharp for the *oct-2* cDNA clone; S. Triezenberg for an HSV stock and C. Preston and S. Triezenberg for advice on preparation of the HSV virion lysate; A. Minson for the LP1 anti-VP16 monoclonal antibody; A. Morris for GH3 pituitary cell RNA; M. Iacovacci for synthesis of oligonucleotides; C. Pabo for the coordinates of the λ repressor/DNA co-crystal structure; J. Duffy and D. Greene for artwork; N. Dyson, M. Giiman, N. Hernandez, B. Stillman, R. Sturm and S. Triezenberg for comments on the manuscript; M. Clearly for editorial corrections; and S. Harrison for helpful discussions. S. S. is the recipient of a NIH postdoctoral fellowship, and W. H. is a Rita Allen Foundation Scholar. This work was supported by grants from the National Cancer Institute and the NSF.

The jets of 3C120

David J. Axon*, S. W. Unger†, A. Pedlar*, E. J. A. Meurs‡§, D. M. Whittle|| & M. J. Ward§

* Nuffield Radio Astronomy Laboratory, Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK
† Royal Greenwich Observatory, Herstmonceux Castle, Hailsham, East Sussex B27 1RP, UK
‡ European Southern Observatory, Karl-Schwarzschild-Strasse-2, D-8046 Garching bei Munchen, FRG
§ Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 OHA, UK
|| Department of Astronomy, University of Virginia, PO Box 3818, University Station, Charlottesville, VA 22903, USA

CORE-DOMINATED radio sources associated with quasars are a manifestation of the most extreme form of activity in galactic nuclei. In general, the morphology of their inner radio structure is in the form of a jet detected on only one side of the core; the larger-scale radio emission is relatively symmetric[1]. Superluminal motion in some sources[2,3] has led to the suggestion that the ejection of radio-emitting material is relativistic and intrinsically two-sided. The apparent one-sidedness of the jets is then explained by relativistic aberration[4–6]. This persuasive interpretation has not escaped criticism: both physical[7] and statistical[8] arguments have been advanced in favour of one-sided ejection. However, our new optical observations of 3C120, which reveal the details of the interaction between the radio jet and the quiescent gas in the galaxy, offer significant kinematic evidence in favour of the relativistic-beaming hypothesis.

The radio source 3C120, at a redshift of 0.033 (ref. 9), is the nearest radio source with a one-sided jet[10,11] that exhibits superluminal motion[12]. Previous low-spatial-resolution spectroscopy[9,13] of the nebulosity surrounding 3C120 revealed underlying rotational motion in the gas, punctuated by large chaotic perturbations.

Recently a number of Seyfert galaxies have been found that have high-velocity narrow-line-region components that are spatially associated with their off-nuclear radio structure[14–16]. It is thought that these high-velocity components result from the interaction between ejected radio-emitting material and ambient gas in the surrounding galaxy[17–19]. In this process the gas experiences a strong shock, which both compresses and accelerates it[14,19]. The high-velocity emission-line components originate in cooling post-shock material that has been photoionized by ultraviolet radiation from the nucleus. To determine whether the complex motions observed in 3C120 could, as in the Seyferts, be of a systematic nature, we have re-investigated the velocity field at high spatial and spectral resolution. Here we show that close to the radio jet, far from being chaotic, the motions are a direct consequence of the interaction of the jet with the interstellar medium in the galaxy. This interaction creates an expanding cocoon of material around the jet which, in narrow-band images, appears as an emission-line counterpart to the radio jet[20].

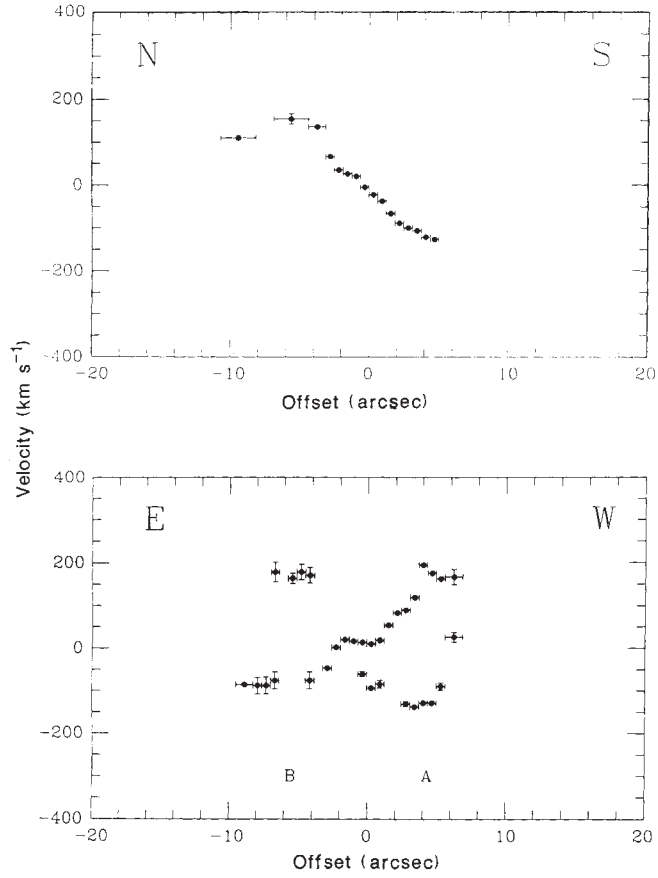


FIG. 2 The heliocentric radial velocities derived from [O III], $\lambda = 5{,}007$ Å for the spectra along PA = 359° (top) (passing through the nucleus), and along PA = 269° (bottom) (along the jet). The velocities shown are relative to 9,920 km s^{-1}, the systemic velocity of 3C120. Integration times of 10,000 and 16,886 s were used for PA = 359° and PA = 269° respectively, and the seeing was 1 arcsec. The position of the centre of the galaxy was determined by fitting a Cauchy function to the nuclear continuum.

We observed 3C120 at several epochs between 1985 December 21–25 and 1988 November 21–25, using the Intermediate Dispersion Spectrograph at the $f/15$ Cassegrain focus of the 2.5-m Isaac Newton Telescope and the Image Photon Counting System[21]. The slit positions discussed here are shown in Fig. 1, superimposed on a 1,666-MHz radio map of the inner jet structure of 3C120 obtained with the MERLIN[22] interferometer.

The variations of radial velocity along position angle 359°, the approximate major axis of the stellar light distribution[20], is

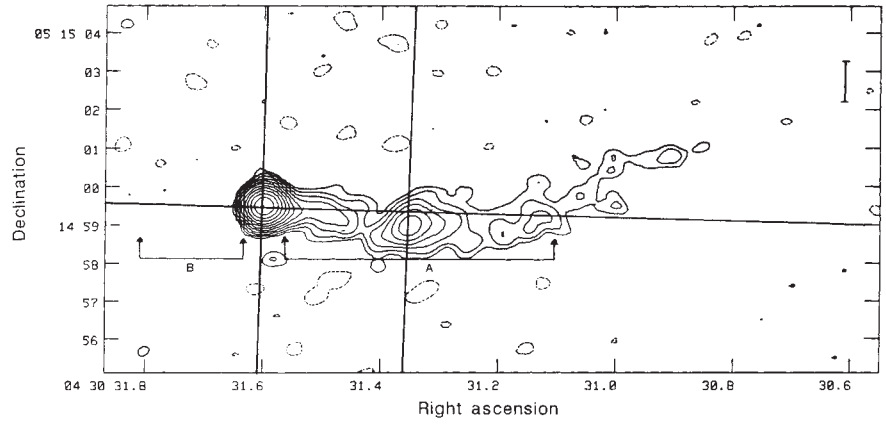


FIG. 1 A 1660-MHz map of 3C120 made with the MERLIN radio telescope. The contour levels are −2.0, −1.0, 1.0, 2.0, 4.0, 8.0, 16.0, 32.0, 64.0, 128.0, 256.0, 512.0, 1024.0 and 2048.0 mJy per beam and the peak flux density in the map is 2.06 Jy per beam. The slit positions discussed in the text are marked with solid lines and the FWHM of the seeing during the observations is shown by the vertical bar. All the spectra have a pixel size of 0.63 arcsec in the spatial direction and cover the spectral lines of Hβ, $\lambda = 4{,}861$ Å and [O III], $\lambda = 4{,}959$ Å, 5,007 Å at a resolution of 0.45 Å. A slit length of 78 arcsec and a width of 0.63 arcsec was used for all observations. The regions where two velocity components are observed are identified here and in subsequent figures by the letters A and B.

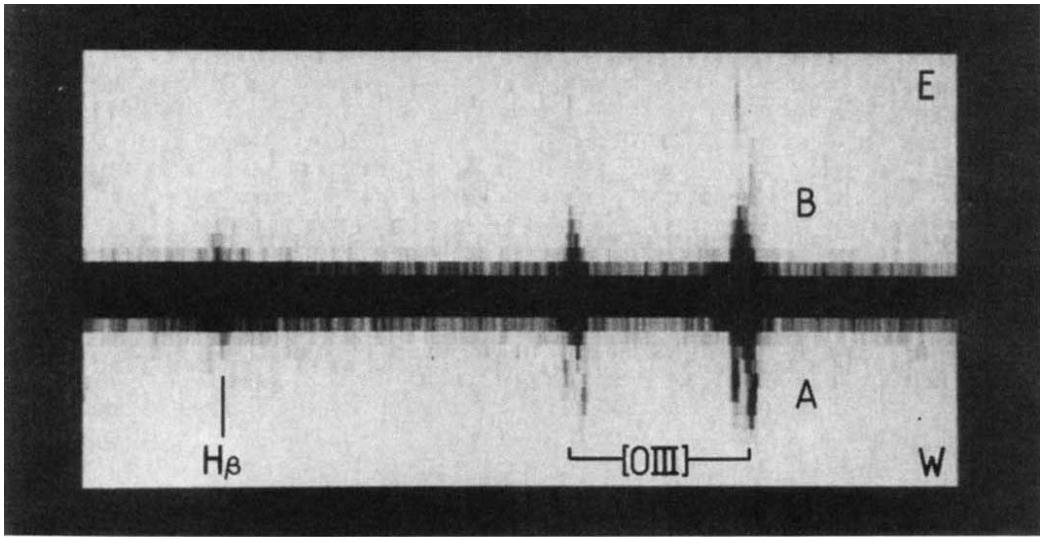


FIG. 3 A grey-scale image of the long-slit spectrum in PA = 269° (along the jet) resulting from a 12,000-s integration. The wavelength scale increases to the right and the Hβ and [O III] emission lines are identified. The orientation is as marked. The FWHM of both velocity components are very similar along the length of the jet and vary between 100 and 150 km s^{-1}. The pixels in this image do not correspond to true detector pixels.

shown in Fig. 2*a*. Over the central 8 arcsec the velocity varies approximately linearly with position, with an amplitude of ±180 km s^{-1}, but at larger radii it flattens and eventually starts to turn over. Such behaviour is characteristic of the rotation curves of most disk galaxies[23]. This picture of 3C120 as a rotating disk galaxy is supported by a more detailed spectroscopic study which we will describe elsewhere.

The spectrum along the radio jet (position angle 269°) is shown in Fig. 3. To the west of the nucleus in region A, which is coincident with the radio jet, it is immediately apparent that the lines are split into two components. This splitting occurs continuously along the bright inner 6–7 arcsec of the radio jet. We measure the maximum separation of the blueshifted and redshifted components to be ~300 km s^{-1} (Fig. 2, bottom). At ~7 arcsec from the nucleus, the radio jet starts to bend to the north-west, and thus out of our slit. The [O III] emission stops abruptly at this point, suggesting that it is spatially associated with the jet. The observation perpendicular to the jet (Fig. 4) greatly strengthens this association, as it shows that the split-line region is localized to the jet and also that the surface brightness of [O III] is enhanced by a factor of 2–3 on the jet. The full width at half maximum (FWHM) of this emission peak is 3.2 ± 0.2 arcsec. Assuming that the Hubble constant H_0 is 75 km s^{-1} Mpc^{-1} then, at the distance of 3C120, 1 arcsec = 643 pc so the diameter of the split-line region is ~2 kpc.

The situation in 3C120 seems physically similar to that in the Seyfert galaxies, but with one subtle distinction: the median velocity of region A is within 30 km s^{-1} of the systemic velocity of the galaxy (9,920 km s^{-1}), suggesting that the split lines do not originate in material that has been ejected with the radio-emitting plasma, but arise instead because of the interaction, most probably lateral expansion, of the radio jet and the interstellar medium of the galaxy[17,18].

If we assume that the jet is in the plane of the sky, then the measured expansion velocity of 150 km s^{-1} implies that the dynamical timescale for the formation of the split-line region is ~6×10^6 yr. This timescale is considerably shorter than the best estimates of the age of the radio lobes[10], and suggests that we are witnessing an expansion that is due to a recent outburst of the jet. The apparent width of the radio jet increases linearly with distance from the core[10]. Hence, for the jet to maintain its identity, the measured width-to-length ratio of ~0.1, coupled with this estimate for the expansion velocity, implies a bulk velocity for the jet material in excess of ~1,500 km s^{-1}. If, however, the jet makes an angle of 12–24° to the line of sight, as implied by the superluminal motion in the core[12], de-projection increases the lateral expansion velocity to ~350–700 km s^{-1}, leading to minimum ejection velocities in excess of 9,000–35,000 km s^{-1}. An independent limit on the angle of the jet to the line of sight of ≥10° is provided by the size of the emission-line region in conjunction with cooling-length arguments[19], as the shocked gas must have cooled to produce the observed emission peak.

The most intriguing aspect of our results is that velocity splitting, of comparable magnitude, is also seen to the east of the nucleus in region B, where no radio jet is detected. By analogy with region A, we believe that the natural explanation for the velocity structure of region B is that it is due to the interaction between a relativistic jet pointing away from us, and ambient gas in the galaxy. As the emission lines originate from material that is not moving relativistically, their intensity is not

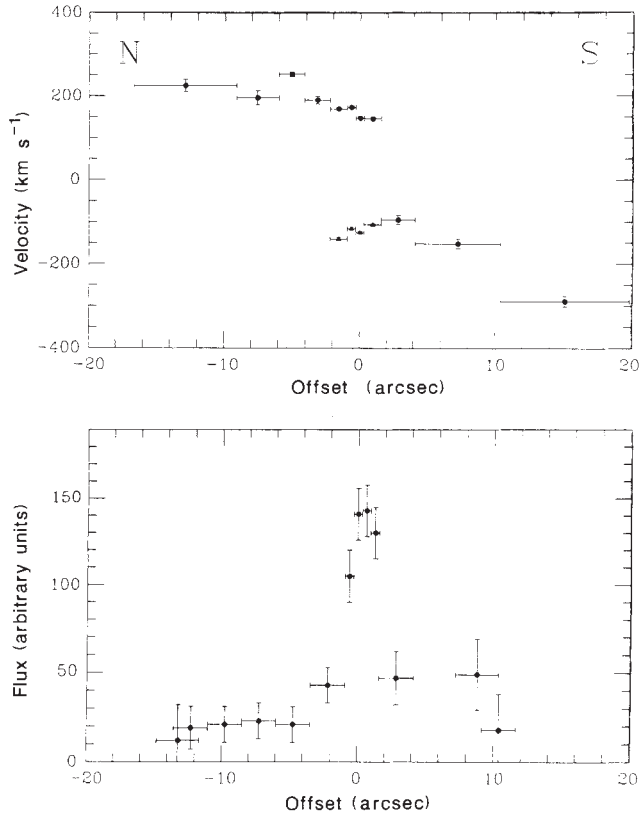


FIG. 4 The heliocentric radial velocities (top) and intensities (bottom) derived from the spectrum along PA = 359° and centred 4 arcsec west of the nucleus. Zero offset marks the centre of the radio jet. Where the emission lines are double, the intensities in the bottom figure are the sums of both components. The total integration time was 15,865 s and the seeing was 1.5 arcsec.

directly dependent on the orientation of the jet, whereas the intensity of the jet itself will be greatly diminished by relativistic aberration. As region B is further away from us its decreased surface brightness in [O III], compared with region A, would be the result of a combination of increased extinction and the fact that it has less cooled gas (the light travel time from region B makes it younger than region A by about 4×10^5 yr).

Can we rule out the alternative explanation that region B is a relic from a previous epoch when a single non-relativistic jet pointed in what is now the counter jet direction? A strong argument against this interpretation is that the expansion velocities on both sides of the nucleus are rather similar. To zero order, the stopping time of the shell is roughly the same as the acceleration time so this requires that region B was created no more than $\sim6\times10^6$ yr ago if it is not being driven by a jet now. Even assuming that the western jet is in the plane of the sky, its measured length, in conjunction with this timescale for its creation, would necessitate a single jet to be travelling at velocities close to the velocity of light.

It is clear from Fig. 4 (top) that the split-line region does not form a closed velocity loop as would be expected from a shell expanding in a uniform medium. However, hydrodynamic models[24] of expanding bubbles in a stratified medium, of the sort expected in disk galaxies, show that once the shell radius exceeds the scale height of the disk it will expand preferentially perpendicular to the plane of the disk because of the reduction in pressure. Commensurate with this 'puncturing' of the bubble, there will also be an increase in the cooling time and a decrease in the emission measure of the material forming the top and bottom of the shell. Consequently, only those parts of the shell that are 'snow-ploughing' through the disk will contribute to the observed velocity distribution. As the present radius of the shell in 3C120 is ~1 kpc, the velocity behaviour in Fig. 4 can be explained in this way provided that 3C120 is a disk galaxy with a typical scale height of ~0.5 kpc.

Observations with the Hubble Space Telescope would enable the kinematics of the gas to be investigated in the inner regions of the jet of 3C120 and in similar more distant objects. The most exciting prospect, however, would be to trace the splitting out to both radio lobes, thereby indicating a currently open fuel link with the central powerhouse. It would indeed be ironic if the controversy surrounding the origin of one-sided radio jets was resolved not by studying the jets themselves but by observing the interaction between the jets and their environment at optical wavelengths. □

1. Bridle, A. H. & Perley, R. A. *A. Rev. Astr. Astrophys.* **22,** 319–358 (1984).
2. Schilizzi, R. T. & de Bruyn, A. G. *Nature* **303,** 26–31 (1983).
3. Zensus, J. A. & Pearson, T. S. (eds) *Superluminal Radio Sources* (Cambridge University Press, 1987).
4. Browne, I. W. A. *et al. Nature* **299,** 788–793 (1982).
5. Scheuer, P. A. G. & Readhead, A. C. S. *Nature* **277,** 182–185 (1979).
6. Blandford, R. D. & Königl, A. *Astrophys. J.* **232,** 34–48 (1979).
7. Rudnick, L. & Edgar, B. K. *Astrophys. J.* **279,** 74–85 (1984).
8. Wardle, J. F. C. & Potash, R. I. in *NRAO Workshop 9 on Physics of Energy Transport in Extragalactic Radio Sources* (eds Bridle, A. H. & Eilek, J. A.), 20–34 (NFAO, Green Bank, 1984).
9. Baldwin, J. A., Carswell, R. F., Wampler, E. J., Burbidge, E. M. & Boksenberg, A. *Astrophys. J.* **236,** 388–425 (1980).
10. Walker, R. C., Benson, J. M. & Unwin, S. C. *Astrophys. J.* **316,** 546–572 (1987).
11. Benson, J. M. *et al. Astrophys. J.* **334,** 560–572 (1988).
12. Walker, R. C., Benson, J. M., Seielstad, G. A. & Unwin, S. C. in *IAU Symp. 110: VLBI and Compact Radio Sources* 121–124 (Reidel, Dordrecht, 1984).
13. Moles, M., del Olmo, A., Masegosa, J. & Perea, J. D. *Astr. Astrophys.* **197,** 1–13 (1988).
14. Whittle, D. M. *et al. Mon. Not. R. astr. Soc.* **221,** 189–200 (1985).
15. Whittle, D. M. *et al. Astrophys. J.* **326,** 125–145 (1988).
16. Pedlar, A. *et al. Mon. Not. R. astr. Soc.* **238,** 863–880 (1989).
17. Pedlar, A., Dyson, J. E. & Unger, S. W. *Mon. Not. R. astr. Soc.* **214,** 463–473 (1985).
18. Pedlar, A., Unger, S. W., Dyson, J. E. & Axon, D. J. *Star Formation in Galaxies NASA Conf. publ. 2466* 711–713 (NASA, Pasadena, 1987).
19. Taylor, D., Dyson, J. E., Axon, D. J. & Pedlar, A. *Mon. Not. R. astr. Soc.* **240,** 487–500 (1989).
20. Hua, C. T. *Astr. Astrophys.* **199,** 105–113 (1988).
21. Boksenberg, A. & Burgess, V. *Adv. Electron. Electron Phys.* **33B,** 835–849 (1972).
22. Davies, J. G., Anderson, B. & Morrison, I. *Nature* **288,** 64–66 (1980).
23. Rubin, V. C., Ford, W. K. Jr & Thonnard, N. *Astrophys. J.* **238,** 471–487 (1980).
24. Tomisaka, K. & Ikeuchi, S. *Publs astr. Soc. Japan* **34** 461–481 (1982).

ACKNOWLEDGEMENTS. We thank J. Dyson, R. Carswell, C. Walker and I. Browne for discussions, the Panel for the Allocation of Telescope Time for telescope time and T. Muxlow for the MERLIN radio map of 3C120. P. Wilkinson helped with the March 1987 observation. E.J.A.M. was supported in part by an ESA research fellowship.

Is the sub-millisecond pulsar strange?

Joshua A. Frieman & Angela V. Olinto

NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory, PO Box 500, Batavia, Illinois 60510, USA

THE interpretation of 1,968.629-Hz optical pulsations[1] from supernova 1987A as a rapidly rotating neutron star poses a challenge to models of nuclear matter. The simultaneous requirement that a star with angular velocity $\Omega_{SN}=1.237\times10^4\ s^{-1}$ be stable against non-axisymmetric perturbations and that the maximum mass of a non-rotating neutron star exceed 1.44 $M_\odot$ (the mass of the binary pulsar) seems to rule out all standard equations of state[2,3]. Three-flavour (up, down, and strange) quark matter, or 'strange matter', may be absolutely stable[4,5], that is, it may comprise the ground state of the hadrons. Neutron stars may become strange stars[6–8], which have a different equation of state and mass–radius relation from conventional neutron-star models. For a range of hadron parameters, the maximum rotation rate of secularly stable strange stars may exceed that of the half-millisecond pulsar and the non-rotating maximum mass M_{max} is greater than 1.52 $M_\odot$. The low-mass companion(s) to SN1987A, inferred from the periodic modulations of the optical signal, can be accounted for by stable strange-matter lump(s) ejected from the young strange star.

Once a stable seed of strange matter is formed inside a neutron star, it grows without bound and converts the star to strange matter[6–8]. Strange stars are thus fundamentally distinguished from neutron stars with quark cores by the assumption that quark matter is stable down to zero pressure. In the context of a bag-model calculation, this assumption has been shown to be plausible[4] for a range of values of the strange-quark mass m_s, the bag constant B (the difference in energy density between the perturbative vacuum and the true vacuum) and the strong coupling constant α_c. (In Fig. 1, we show contours of fixed energy per baryon number for strange matter, $E/A=899$ and 939 MeV, in the limit $\alpha_c=0$.) Although these three parameters have been estimated by fitting the bag model to the masses of light hadrons, these fits introduce additional variables, making it difficult to extract the values appropriate for bulk quark matter[4]. Also, the parameters m_s and α_c drop with increasing density, so fits made to low-density data may not be adequate for assessing strange-matter stability in neutron stars. (Furthermore, the bag model is probably a crude approximation to the behaviour of bulk quark matter; more appropriate might be a model in which B is a dynamical variable, for example an $SU(3)\times SU(3)$ sigma model (B. Lynn, private communication).) In exploring stellar models based on strange matter, we shall therefore impose only the self-consistency criterion, choosing parameters that lie in the estimated window of strange-matter stability, $E/A<939$ MeV. As one moves away from the stable strange-matter parameter region, neutron stars have shrinking quark cores, surrounded by growing envelopes of neutron matter. In this case, the transition to neutron matter typically takes place at or below nuclear density, $\rho_{tr}\lesssim\rho_{nuc}=2.8\times10^{14}\ g\ cm^{-3}$, if $E/A\lesssim950$–960 MeV (depending on the neutron equation of state). For the corresponding hadron parameter range, the envelope has a negligible role in the structure of the star and our results below carry through unchanged.

The equation of state (EOS) for strange matter can be written

$$p=\tfrac{1}{3}(\rho-4B)-\tfrac{1}{3}(\rho-B)\varepsilon. \qquad (1)$$

The function $\varepsilon(m_s(4B)^{-1/4},\rho/4B)<0.14$ is a measure of the correction to the equation of state that accounts for the non-zero m_s and vanishes in the limits of $m_s=0$ (where strange (s), up (u) and down (d) quarks have equal abundances) and $m_s\simeq1.6(4B)^{1/4}$ (because the s-quark abundance is suppressed as m_s

approaches μ_s, the chemical potential). In calculating ε, we neglect $O(\alpha_c)$ corrections that result from quark interactions within the bag. These effects are small for the EOS, but they shift the window of strange-matter stability to lower values of B (ref. 4). We have constructed spherical general-relativistic star models using this equation of state. We note that for an EOS of the form of equation (1), the QCD energy scale $B^{1/4}$ can be scaled out of the Oppenheimer-Volkoff equation of hydrostatic equilibrium. Mass-radius sequences are thus characterized by the single dimensionless ratio $m_s/(4B)^{1/4}$. A number of consequences follow from this scale invariance. For example, for fixed $m_s/(4B)^{1/4}$, the ratio of the stellar radius to the Schwarzschild radius for maximum-mass models, $R_{max}/2GM_{max}$, is independent of B.

Parameters of the maximum-mass models are given in Table 1 in terms of a fiducial bag constant $B_0^{1/4}=145$ MeV ($B_0=57$ MeV fm^{-3}). Because the mass scales as $M\sim B^{-1/2}$, the maximum mass reaches a lower bound at the edge of the strange-matter stability window, $E/A=939$ MeV; there, $M_{max}>1.52\,M_\odot$, with $R_{max}>8.52$ km. By contrast with some soft equations of state, this is large enough to account for the observed mass[9] of PSR1913+16 and is consistent with other inferred binary masses; neutron-star mass observations to date place no constraints on the allowed parameter space.

The maximum rotation rates of neutron stars depend sensitively on the mass distribution and thereby on the equation of state. Generally, soft models are more centrally condensed, have smaller radii and maximum masses and can rotate faster than stiff models. Strange matter embodies elements of both soft and hard equations of state. For example, at the low-mass end, where gravity plays essentially no role in their structure, strange stars are approximately uniform-density spheres with $\rho\simeq 4B$ and $M\sim R^3$. Low-mass strange stars, and the outer regions of strange stars that are more massive, are much stiffer than all neutron-star models. In the central regions of strange stars near the maximum mass, however, $\rho_c\simeq 5(4B)$ and the equation of state gets very soft (relativistic Fermi gas).

Massive strange stars are thus hard on the outside and soft on the inside. The stiffness is reflected in the relative flatness of the density profile: for maximum-mass models, $\rho_c/\bar\rho=2.1$–2.7, lower than or comparable to the stiffest neutron-star models. On the other hand, for large values of B, the maximum-mass strange stars have global features comparable to soft nuclear models. For example, consider models along the contour of constant radius $R=8.7$ km in Fig. 1. At $m_s=0$, we find $M_{max}=1.59\,M_\odot$, with average density $\bar\rho=1.15\times10^{15}$ g cm^{-3} and moment of inertia $I=1.15\times10^{45}$ g cm^2. As one moves up to $m_s/(4B)^{1/4}=0.75$, these values drop by less than 3%, whereas the central density decreases from 3.09×10^{15} to 2.75×10^{15} (the density profile flattens). Of the models in the Arnett-Bowers collection[10], this is closest to EOS A, the Reid soft core[11]. Also, at $m_s=0$, for the maximum-mass strange star, the pressure-averaged adiabatic index $\bar\Gamma=\bar\rho/(\bar\rho-4B)$ is 2.3, which is moderately soft.

How fast can strange stars spin? A lower bound on the maximum rotation rate is given by the low-mass models with $m_s=0$. Although low-mass strange stars are bound principally by the strong interactions, not gravity, their stability against rotationally induced fission into smaller fragments is determined by gravity because the gravitational energy increases when the star is deformed by rotation, but the nuclear binding energy does not. For $M<0.5(B/B_0)^{-1/2}M_\odot$, the ratio of central to average density $\rho_c/\bar\rho<1.06$, the pressure $p_c<0.06\rho_c$, the average polytropic index $\bar n=(\bar\Gamma-1)^{-1}<0.1$, and the post-newtonian corrections to gravity are negligible. When they rotate these ultra-stiff low-mass stars are excellent approximations to newtonian Maclaurin spheroids (the Maclaurin approximation improves as the mass decreases). As the average density is bounded from below by $\bar\rho>4B$, the maximum rotation rate, which is determined by the dynamical instability[12], is $\Omega_{dyn}=0.67(\pi G\bar\rho)^{1/2}>6.22\times10^3(B/B_0)^{1/2}$ s^{-1}. The non-axisymmetric secular $m=2$ instability, driven by gravitational radiation reaction (GRR)[13], sets in at $\Omega_{sec}=0.91\Omega_{dyn}$. Although there are also modes with $m>2$ that become unstable at lower angular velocity, strange stars can rotate fast enough to account for the 1.6-ms pulsar.

More-massive strange stars can rotate faster because they have higher densities. For strange stars near the maximum mass, we estimate the maximum rotation speed using the results[2,3,14] for other equations of state. An absolute upper limit is set by the onset of equatorial (keplerian) mass shedding. For all the models considered in refs 2, 3, and 14, the keplerian limit of the maximum-mass star was found to lie in the range $\Omega_k=(0.76\pm0.04)(\pi G\bar\rho_s)^{1/2}$, where $\bar\rho_s$ is the average density of the spherical (non-rotating) maximum-mass model. To account for the apparent angular velocity of PSR1987A, $\Omega_k>1.237\times10^4$ s^{-1}, the density must satisfy the constraint $\bar\rho_s>(1.12-1.37)\times10^{15}$ g cm^{-3}. Maximum-mass strange stars with average density in this range correspond to the parameter region between the two dashed curves of Fig. 1. Thus, within the uncertainties, there is a range of hadronic parameters for which strange stars may rotate fast enough to account for the 0.5-ms pulsar. Clearly, a more accurate determination requires rapidly rotating strange-star models.

Although some conventional neutron-star models, such as those based on EOS A or F[2], also satisfy the constraint $\Omega_k>\Omega_{SN}$, they are secularly unstable to perturbations in the $m=4$ or 5 mode[2,3,15–19]. Because these GRR-driven instabilities are damped by viscosity, the origin of the difficulty is the relatively low viscosity of hot young neutron matter. In this regard, strange matter offers two advantages. First, at equal temperature and density, the shear viscosity of strange matter[20] $\eta_s\approx 9.7\times10^{16}(\alpha_c/0.13)^{-3/2}(\rho/10^{15})^{5/3}(T/10^{10}\,\mathrm{K})^{-2}$ g cm^{-1} s^{-1}, is roughly an order of magnitude larger than that of nuclear matter[21], $\eta_{nuc}\approx 9.6\times10^{15}(\rho/10^{15})^{9/4}(T/10^{10}\,\mathrm{K})^{-2}$. Here, the strange-matter viscosity has been calculated to lowest order in α_c, assuming the magnetic field $B\leqslant10^{10}(T/10^9\,\mathrm{K})^2$ G. (This is consistent with the observatonal upper bound on the field of SN1987A determined by the continued decline in bolometric luminosity.) Second, neutrino cooling in strange stars is significantly more efficient than the modified URCA process[22]. As a result, the core temperature of a two-year-old strange star is estimated to be $\sim10^8$ K, as opposed to $\sim10^9$ K for neutron stars. As the shear viscosity scales as T^{-2}, the present strange-star viscosity, $\bar\nu\simeq6\times10^5$ cm^2 s^{-1}, is a factor $\sim10^3$ larger than that for young neutron stars. Consequently, GRR-driven secular instabilities are more effectively damped in strange stars. Using the estimated viscous-damping timescale[15–17,19], we find that the $m=3$ mode would limit the current angular velocity of the

TABLE 1 Maximum-mass models

$m_s(4B)^{-1/4}$	$M_{max}(B/B_0)^{1/2}(M_\odot)$	$R_{max}(B/B_0)^{1/2}$ (km)	$\rho_c(B/B_0)^{-1}$ (g cm^{-3})	$\bar\rho(B/B_0)^{-1}$	$R_m/2GM_m$
0	2.006	10.95	1.95×10^{15}	7.27×10^{14}	1.847
0.25	1.973	10.80	1.97×10^{15}	7.44×10^{14}	1.854
0.5	1.887	10.43	2.02×10^{15}	7.90×10^{14}	1.871
0.75	1.775	9.94	2.11×10^{15}	8.58×10^{14}	1.897
1.0	1.665	9.54	2.09×10^{15}	9.10×10^{14}	1.940
1.2	1.594	9.27	2.05×10^{15}	9.52×10^{14}	1.968

Fiducial bag constant $B_0^{1/4}=145$ MeV.

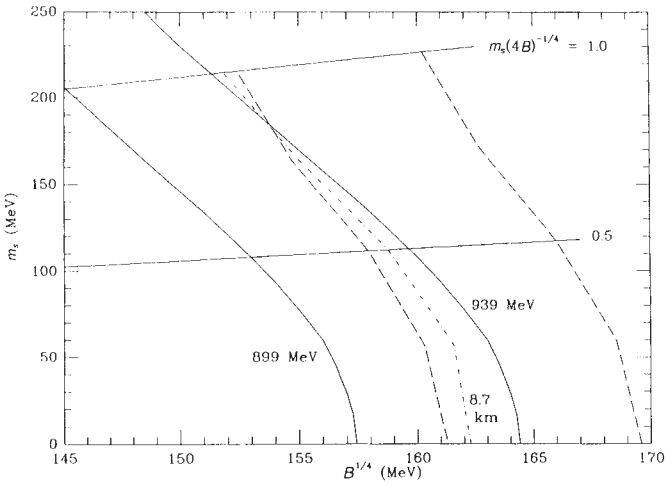


FIG. 1 The strange-matter parameter space. Solid curves marked 899 and 939 MeV are contours of constant energy per baryon number for bulk strange matter (from ref. 4). Dot-dashed curve is the contour of constant radius $R=8.7$ km for non-rotating maximum-mass models. The area between the dashed curves indicates the region within which the expected keplerian rotation speed $\Omega_k=\Omega_{SN1987A}$. Also shown are lines of constant $m_s/(4B)^{1/4}=$ 1.0 and 0.5.

putative strange pulsar. For this mode and for the viscosity above, the pulsar rapidly spins down if $\Omega \geqslant (0.98\text{-}0.99)\Omega_k$ (refs 16, 17, 23). Thus, the secular instability does not substantially narrow the possibilities for rapidly rotating strange stars. By contrast, a two-year-old neutron star is unstable, on a timescale shorter than the observed lower limit for SN1987A, if $\Omega \geqslant (0.92\text{-}094)\Omega_k$ (ref. 23). Also, at earlier times (and therefore higher temperatures), the bulk viscosity of strange matter can be quite large[24] (depending on m_s) and may have a role in stabilizing the secular modes.

The reported 8-hour modulation of the optical pulsar period has been interpreted as evidence for a Jupiter-mass companion to SN1987A (ref. 1). Unlike neutron stars, strange stars are bound by the strong force rather than gravity; they have no minimum stable mass. Thus, planetary-mass strange-matter debris ejected in an asymmetric supernova collapse, or shed in a rapidly rotating post-collapse phase, would not disintegrate. Furthermore, strange-matter lumps have a density comparable to that of strange stars, so they can be ejected without being disrupted by tidal forces.

The SN1987A pulsar data also seems to have a second amplitude and phase modulation, with a period of ≈ 2 h. If this modulation is attributed to a second companion, it implies a Neptune-mass planet on a highly eccentric orbit with semi-major axis of about 6.5×10^5 km (J. Middleditch, personal communication). To survive tidal disruption, such a satellite must have a mean density $\bar{\rho} \geqslant 25$ g cm^{-3}. This is much higher than the density of any planet and the inferred mass is well below the lower mass limit for white dwarfs. Only a strange-matter lump seems to be consistent with this. An alternative explanation is modulation that arises from the precession of the pulsar; this hypothesis could be tested with further data. If precession is ruled out, the case for strange matter would be compelling. □

Received 9 June; accepted 5 September 1989.

1. Kristian, J. *et al. Nature* **338**, 234-236 (1989).
2. Friedman, J. L., Ipser, J. & Parker, L. *Astrophys. J.* **304**, 115-139 (1986).
3. Friedman, J. L., Ipser, J. & Parker, L. *Phys. Rev. Lett.* **62**, 3015-3019 (1989).
4. Farhi, E. & Jaffe, R. L. *Phys. Rev.* D **30**, 2379-2390 (1984).
5. Witten, E. *Phys. Rev.* D **30**, 272-285 (1984).
6. Alcock, C., Farhi, E. & Olinto, A. V. *Astrophys. J.* **310**, 261-272 (1986).
7. Olinto, A. V. *Phys. Lett.* B **192**, 71-75 (1987).
8. Haensel, P., Zdunik, J. & Schaeffer, R. *Astr. Astrophys.* **160**, 121-128 (1986).
9. Weisberg, J. & Taylor, J. *Phys. Rev. Lett.* **52**, 1348-1351 (1984).
10. Arnett, W. D. & Bowers, R. L. *Astrophys. J. Suppl.* **33**, 415-436 (1977).
11. Pandharipande, V. *Nucl. Phys.* A**174**, 641-656 (1971).
12. Chandrasekhar, S. *Ellipsoidal Figures of Equilibrium* (Yale University Press, 1969).
13. Chandrasekhar, S. *Astrophys. J.* **161**, 561-569 (1970).
14. Butterworth, E. & Ipser, J. *Astrophys. J.* **204**, 200-223 (1976).
15. Friedman, J. L. *Phys. Rev. Lett.* **51**, 11-14 (1983).
16. Lindblom, L. *Astrophys. J.* **303**, 146-153 (1986).
17. Cutler, C. & Lindblom, L. *Astrophys. J.* **314**, 234-241 (1987).
18. Imamura, J., Friedman, J. L. & Durisen, R. *Astrophys. J.* **294**, 474-478 (1985).
19. Managan, R. *Astrophys. J.* **294**, 463-473 (1985).
20. Haensel, P. & Jerzak, A. *Acta Phys. Pol.* B **20**, 141-161 (1989).
21. Flowers, E. & Itoh, N. *Astrophys. J.* **230**, 847-858 (1979).
22. Alcock, C. & Olinto, A. *A. Rev. Nucl. Particle Sci.* **38**, 161-184 (1988).
23. Ipser, J. & Lindblom, L. *Phys. Rev. Lett.* **62**, 2777-2780 (1989).
24. Sawyer, R. *Phys. Rev.* D **39**, 3804-3806 (1989).

ACKNOWLEDGEMENTS. We thank F. Grassi, J. Middleditch and R. Sawyer for discussions. This work was supported in part by U.S. Department of Energy and by NASA at Fermilab.

Tropospheric lifetimes of halogenated anaesthetics

A. C. Brown*, C. E. Canosa-Mas*, A. D. Parr*, J. M. T. Pierce† & R. P. Wayne*

* Physical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK
† Shackleton Department of Anaesthetics, Southampton General Hospital, Southampton SO9 4XY, UK

CONCERN has been expressed over the use of the halogenated anaesthetics halothane ($CF_3CClBrH$), enflurane (CF_2HOCF_2CFClH) and isoflurane ($CF_2HOCHClCF_3$) because of their potential for stratospheric ozone destruction[1]. Halogenated species also contribute to global warming[2]. The significance of the anaesthetics in stratospheric ozone loss or in 'greenhouse' heating depends on their atmospheric lifetimes. Because reaction with hydroxyl (OH) radicals is likely to be the main homogeneous sink for these species in the troposphere, we have measured absolute rates of reaction with OH. Comparison with a one-dimensional model[3] indicates that the lifetimes of halothane, enflurane and isoflurane with respect to this reaction are 2, 6 and 5 years, respectively. Thus the small production of the anaesthetics is not offset by anomalously long atmospheric lifetimes to give a large atmospheric burden of the compounds. The anaesthetics will contribute at most a fraction of $\sim5\times10^{-4}$ to the total atmospheric content of chlorine-containing species.

Halothane, enflurane and isoflurane are popular potent inhalation anaesthetics[4]. Almost all of the administered doses will end up in the atmosphere[5], but only those species with a tropospheric lifetime of more than about two years will reach the stratosphere in significant quantities.

The efficiency with which a unit mass of each anaesthetic will destroy stratospheric ozone relative to the most important ozone-depleting molecules, chlorofluorocarbons (CFCs) 11 and 12 (CFC-11 and CFC-12), will depend on the molecular weight of the anaesthetic, the number and type of halogen atoms per molecule and the compound's atmospheric lifetime[3]. The relative ozone-depletion potential (ODP) is given by

$$\mathrm{ODP}=\frac{\tau_{\text{An-H}}}{\tau_{\text{CFC-11}}}\times\frac{M_{\text{CFC-11}}}{M_{\text{An-H}}}\times\frac{n_{\text{Cl}}+\alpha n_{\text{Br}}}{3} \qquad (1)$$

where An-H represents the parent anaesthetic molecule, and $M_{\text{CFC-11}}$ and $M_{\text{An-H}}$ are the relative molecular masses of CFC-11 and the parent anaesthetic species, respectively. The total number of Cl and Br atoms in the molecule are represented by n_{Cl} and n_{Br}, and α is the efficiency of Br in destroying stratospheric ozone relative to that of Cl. According to current estimates[3], Br may be between 35 and 80 times more potent as a catalyst than Cl, so that the factor α must be taken into account in deriving ODPs.

The most important daytime degradation of the anaesthetic molecules will be[6] reaction with the OH radical

$$\text{An-H}+\text{OH} \rightarrow \text{An}+\text{H}_2\text{O}$$

TABLE 1 Rate constants for reaction with OH, atmospheric lifetimes, and other data for the anaesthetics and common halogenated compounds

Compound	Molecular weight mass	$k_2 \times 10^{14}$ ($cm^3 s^{-1}$)	τ (yr)	Potential ozone depletion efficiency	Integrated absorption cross-section ($cm^{-2} atm^{-1}$)	Greenhouse warming potential	Global production (tonnes yr^{-1})
CFC-11 $CFCl_3$	104.5	$<5.0\times10^{-4}$*	76‡	1.0‡	2,507#	0.39#	350,000**
CFC-12 CF_2Cl_2	121	$<6.0\times10^{-4}$*	140‡	1.0‡	3,404#	1.0#	400,000**
CFC-22 CHF_2Cl	86.5	0.46*	9.0§ 22‖	0.1‡	—	0.07#	72,000‡
Methyl chloroform CH_3CCl_3	133.5	1.2*	3.4§ 8.6‖	0.05‡	—	—	474,000‡
Halon 1211 CF_2ClBr	165.5	—	12.5‡	2.7‡	—	—	3,000‡
Halothane $CF_3CHClBr$	197.4	6.0±0.4†	0.7§ 2‖	0.36¶	1,400¶	0.004¶	1,000††
Enflurane CF_2HOCF_2CFClH	184.5	1.7±0.5†	2.4§ 6‖	0.02¶	4,800¶	0.04¶	220††
Isoflurane $CF_2HOCHClCF_3$	184.5	2.1±0.5†	2.0§ 5‖	0.01¶	3,900¶	0.03¶	800††
Sevoflurane $CFH_2OCH(CF_3)_2$	184	7.3±2.2†	0.6§ 1.4‖	0¶	2,550¶	0.005¶	—

* Ref. 11.
† This work.
‡ Ref. 3.
§ Calculated from $\tau=\{k_2[OH]\}^{-1}$, where $[OH]=7.7\times10^5 cm^{-3}$.
‖ τ scaled by comparison with LLNL one-dimensional calculation[3] for CH_3CCl_3 and CHF_2Cl.
¶ This work, calculated with τ values scaled as above‖.
\# Ref. 13.
** Ref. 15.
†† Ref. 5.

Photolysis is another potential loss process, but the ultraviolet spectra of enflurane[7,8] and isoflurane show absorption onsets at $\lambda < 290$ nm, so that photolysis of these compounds does not occur in the troposphere—only in the stratosphere is there significant photolysis. For the bromine-containing species halothane, the onset of absorption seems[7] to be similar, but there may be a long-wavelength tail, as found for CF_2ClBr. The atmospheric lifetime of the latter compound against photolysis in the troposphere and lower stratosphere[3] is ~12.5 yr, and a similar photolytic lifetime might be expected for halothane. We have found that the reaction with OH is about six times faster, so that the rate of the above reaction will effectively determine the tropospheric lifetime τ (defined as $\{k_2[OH]\}^{-1}$) for all the compounds studied.

Our measurements of the second-order rate constants for the reactions between the OH radical and the anaesthetic reagents halothane, enflurane and isoflurane are given in Table 1. Data are also given for a further anaesthetic agent, sevoflurane ($CF_3CH(CF_3)OCH_2F$). As the molecule does not contain Cl or Br atoms, it does not pose a threat to stratospheric ozone, but it was included in our study as a potential 'greenhouse' gas. We used an absolute discharge-flow technique[9,10] to provide time resolution and resonance fluorescence to measure OH concentrations. The total pressures in the flow tube were ~2 torr; oxygen was not present. It is important to know if the rates of reaction with OH are significantly temperature dependent. By analogy with measured activation energies (E_a) for hydrogen-atom abstraction by OH from other halogenated organic compounds[11], we expected a value of E_a of ~10 kJ mol^{-1}. We obtained activation energies of ~8.6 and ~7.6 kJ mol^{-1} for the reactions with enflurane and sevoflurane by performing measurements at two temperatures (300 K and 422 K). This result is in accordance with expectation and, because the value is relatively low, it is not necessary to measure it more precisely or to determine it at temperatures below ambient. A temperature change from 300 K to 220 K induces a change in the rate by at most a factor of three.

We have estimated the tropospheric lifetime of each anaesthetic in two ways. First, we assume[12] a uniform OH-radical concentration of 7.7×10^5 molecule cm^{-3} and a tropospheric temperature of 300 K. The calculated values of τ and the relative ozone-depletion efficiencies (calculated from equation 1) are listed in Table 1. Second, we have scaled the lifetimes determined[3] in one-dimensional model calculations for methyl chloroform (CH_3CCl_3) and CFC-22 (CHF_2Cl) using the published[11] rate constants for reaction with OH for the latter compounds and our new measurements for the anaesthetics. The lifetime, longer by a factor of ~2.5, that we obtained using the second method probably reflects the more realistic treatment of the altitude dependence of both [OH] and temperature in the one-dimensional model. In assessing the catalytic activity for the halogen atoms released by halothane, $CF_3CHClBr$, we have used the ODP obtained from the Lawrence Livermore National Laboratory (LLNL) model[3] for the analogous compound CF_2ClBr as the basis for comparison in both methods.

Long-lived halogenated organic compounds can also contribute to global warming because of their intense absorptions, especially in the 'atmospheric window' between 800 and 1,200 cm^{-1} in the infrared[2]. The greenhouse warming potential (GHWP) of each anaesthetic can be calculated to a first approximation by comparing the integrated absorption cross-section, S, atmospheric lifetime, and molecular weight with CFC-12, defining the GHWP of CFC-12 to be unity[13] so that

$$\mathrm{GHWP}=\frac{\tau_{\text{An-H}}}{\tau_{\text{CFC-12}}}\times\frac{M_{\text{CFC-12}}}{M_{\text{An-H}}}\times\frac{S_{\text{An-H}}}{S_{\text{CFC-12}}} \qquad (2)$$

The infrared absorption spectra of halothane and enflurane have been reported[14] but we have not found any published values of the integrated absorption cross-sections of the anaesthetics. To estimate the GHWP of halothane, enflurane, isoflurane and sevoflurane we measured their infrared spectra with a commercial infrared spectrophotometer (Perkin-Elmer Type 881, 2.40 cm^{-1} resolution) and estimated their integrated absorption cross-sections in the range 800 to 1,200 cm^{-1} from the area under the peaks in the spectra. Experiments with isoflurane showed that the absorbance was a linear function of pressure between 2 and 7 torr of reactant. We observed no change in the absorption spectrum on adding up to 738 torr of N_2, or on cooling the absorption cell from 303 K to 271 K, so that at the resolution employed, the spectra are applicable to tropospheric conditions. Representative spectra for the region 600 to 1,600 cm^{-1} are shown in Fig. 1; integrated absorption cross-sections in the range 800 to 1,200 cm^{-1} and the corresponding GHWPs are listed in

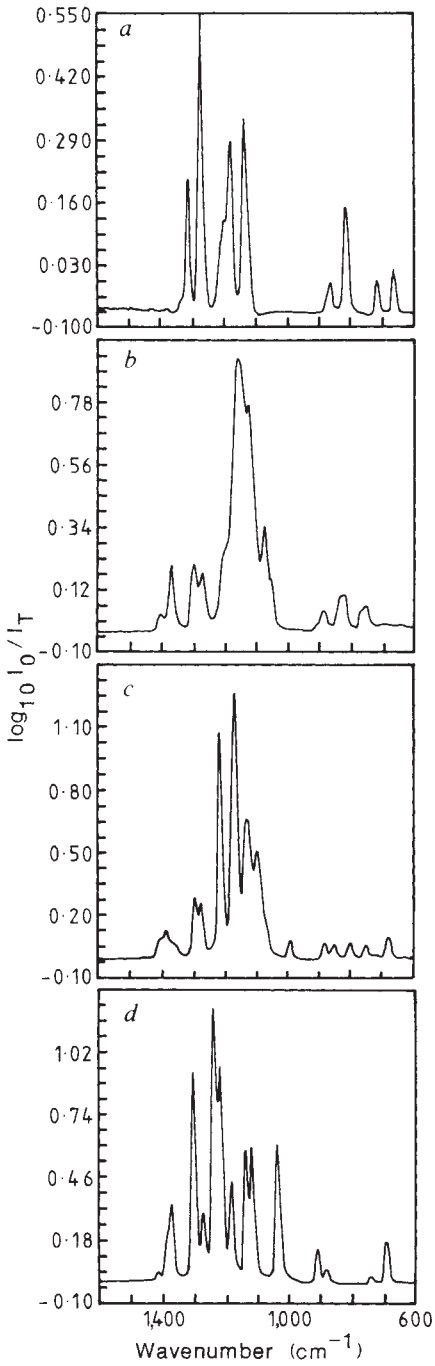


FIG. 1 Infrared spectra of *a*, halothane (2.7 torr); *b*, enflurane (3.1 torr); *c*, isoflurane (3.1 torr); and *d*, sevoflurane (3.1 torr).

Table 1. The most intense features are ether-group stretches in the range 1,000–1,200 cm^{-1}, bands not present in CFCs.

The major uncertainties in our estimate of the tropospheric lifetime of the An-H species are the extent to which the anaesthetics are removed from the atmosphere by further processes such as 'rain-out' and the modelling of global tropospheric OH concentration and temperature. Nevertheless, the relatively short lifetimes of the anaesthetics mean that both the potentials for ozone depletion and the GHWPs are much smaller than for CFC-11 or CFC-12. In addition to the small potentials, the global market for the anaesthetics (1,000, 220 and 800 tonnes for halothane, enflurane and isoflurane[5]) is much less than for CFC-11 and CFC-12 (350,000 and 400,000 tonnes[15]). Thus the overall contribution to ozone depletion or greenhouse warming by the anaesthetic agents must be less than 5 parts in 10^4 of that produced by the CFCs. □

Received 13 July; accepted 11 September 1989.

1. Nørreslet, J. *et al. Lancet* **i,** 719 (1989).
2. WMO Global Ozone Research and Monitoring Project *Atmospheric Ozone Report 16* (World Meteorological Organisation, Geneva, 1985).
3. Hammitt, J. K. *et al. Nature* **330,** 711–716 (1987).
4. Goth, A. *Medical Pharmacology* (C. V. Cosby Co., St Louis, 1981).
5. Rogers, R. C. & Ross, J. A. S. *Lancet* **i,** 1209–1210 (1989).
6. Finlayson-Pitts, B. J. & Pitts, J. N. Jr Atmospheric Chemistry (Wiley Interscience, New York, 1986).
7. Dumas, J.-M. *et al. Can. J. Spectrosc.* **26,** 102–108 (1981).
8. Lewis, B. R., Berzius, L. & Carver, J. M. *Appl. Opt.* **25,** 2647–2648 (1986).
9. Burrows, J. P., Wallington, T. J. & Wayne R. P. *JCS Faraday Trans. II* **79,** 111–122 (1983).
10. Boodaghians, R. B., Hall, I. W. & Wayne, R. P. *JCS Faraday Trans. II* **83,** 529–538 (1987).
11. NASA Panel for Data Evaluation *Chemical Kinetics and Photochemical Data for use in Stratospheric Modeling. No. 8* JPL Publn No. 87–41 (Jet Propulsion Laboratory, Pasadena, 1987).
12. Prinn, R. *et al. Science* **238,** 945–950 (1987).
13. Rogers, J. D. & Stephens, R. D. *J. geophys. Res.* **93,** 2423–2428 (1988).
14. Liu, T. & Ji, L. *Pharm. Ind.* **4,** 37–39 (1984).
15. CMA Fluorocarbon Program Panel *Production, Sales and Calculated Release of CFC11 and CFC12 through 1986* (Chemical Manufacturers Association, Washington, DC, 1987).

ACKNOWLEDGEMENTS. We thank the Meteorological Office for a studentship for A.C.B., the NERC for a grant to support C.E.C.-M., the DoE for a grant adminstered by AERE Harwell for A.D.P., and Michael Kenworthy-Browne and the Maruishi Pharmaceutical Company for samples of the anaesthetics.

Photochemical source of biological substrates in sea water: implications for carbon cycling

David J. Kieber*†, Julie McDaniel*† & Kenneth Mopper*‡

* Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA

DISSOLVED organic carbon (DOC) in sea water represents one of the largest reservoirs of carbon on the earth[1]. The main fraction of this DOC is generally believed to be composed of old[2], biologically refractory material[3] such as humic substances, for which the removal mechanisms remain largely unknown. One potentially important removal process in the ocean that has not been investigated is the photochemical breakdown of this DOC in the photic zone to form biologically labile organic products. Here we show that biological uptake of pyruvate is highly correlated to its rate of photochemical production in sea water ($r = 0.964$), and that the photochemical precursor(s) of pyruvate is from the fraction of DOC having a nominal molecular weight of 500. This is the first evidence that photochemical breakdown of high-molecular-weight marine DOC, which is presumably biologically refractory, results in the production of a compound that is used by plankton as a substrate. Our results have important implications for the oceanic carbon cycle, particularly with respect to planktonic-food-web dynamics and the global carbon budget.

We collected surface seawater samples from a variety of coastal and oligotrophic stations (Fig. 1) during the SOLARS (Study of Light Activated Reactions in the Sea) cruises. Details of sampling procedures will be presented elsewhere (D.J.K. *et al.*, manuscript in preparation). Briefly, we determined the biological uptake of pyruvic acid by measuring its ambient substrate concentration S_n by high-performance liquid chromatography[4] and determining the turnover time t_n of pyruvate at S_n using the respiration-corrected kinetic approach[5]. Typical S_n values ranged from 0.2 to 2.0 nM for oligotrophic stations and 0.9 to 6.2 nM for coastal stations. We added [2-^{14}C]-labelled pyruvate to unfiltered seawater samples to obtain concentrations in the range 10–120 nM; we added glutaraldehyde to pyruvate controls trols to account for abiological processes (such as adsorption

† Present addresses: Chemistry Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA (D.J.K.); Marine Institute, University of Georgia, Sapelo Island, Georgia 31327, USA (J. McD.).
‡ To whom correspondence should be addressed.

onto particles). We incubated the samples in acid-washed polycarbonate bottles placed in a water bath and exposed to sunlight. There is evidence that light inhibits uptake of organic substrates[6]; nevertheless, the incubations were done under light conditions because this most closely reflected ambient conditions in surface sea water. In cases where we made light-dark comparisons to estimate the magnitude of photoinhibition of uptake, the light rates of pyruvate uptake were within 20% of the corresponding dark rates (D.J.K. *et al.*, unpublished results). Samples were incubated from 0.2 to 1.7 h at coastal stations and from 3 to 13 h for offshore sites in the Gulf Stream and the Sargasso Sea. We chose incubation times such that, in most instances, the substrate uptake was <10% at the lowest addition; we measured the amount of label assimilated as well as the amount respired to carbon dioxide. The turnover time is the vertical-axis intercept of the plot of the incubation time divided by the fraction of isotope taken up as a function of concentration of added pyruvate. Time-course experiments performed at each station showed that the uptake of pyruvate in all cases was linear during incubations[7] (D.J.K. *et al.*, manuscript in preparation). After determining S_n and t_n, we can calculate the uptake rate of pyruvate V_n at the ambient substrate concentration from the equation:

$$V_n = S_n / t_n$$

Concurrent with uptake studies, we exposed 0.22-μm-filtered seawater samples to sunlight to determine rates of photochemical production of pyruvate. Irradiations, ranging from 1 to 3 h, were carried out at midday between 10:00 and 14:00 local time. Sample filtration has no significant effect on photochemical production rates of pyruvate when performed carefully (such as gravity filtration through a Whatman GF/C filter followed by vacuum filtration through a 0.22-μm nylon filter at a pressure differential ≤100 mm Hg)[4].

Using data from several field locations, we plotted the uptake rate of pyruvate at S_n against its rate of photochemical production (Fig. 2). Linear regression analysis of the data was performed using a method[8] in which it is assumed that uncertainty is associated with the determination of both variables. The results of this analysis show that uptake rates of pyruvate are

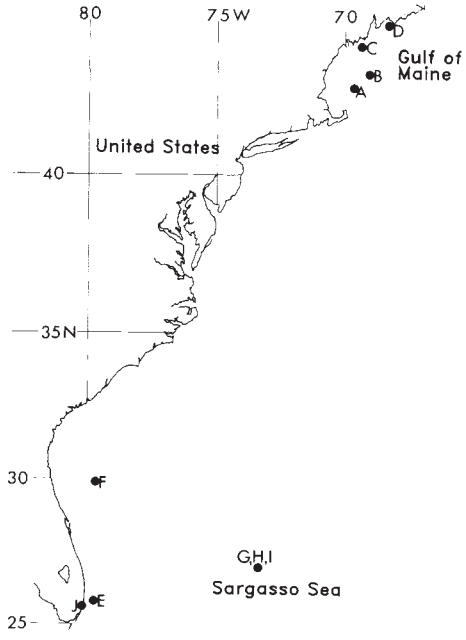


FIG. 1 Map of sampling locations. Gulf of Maine: *a*, Wilkinson Basin, 42°35.2′ N, 69°34.9′ W; *b*, Ammen Rock, 42°53.8′ N, 68°56.6′ W; *c*, Monhegan Island, 43°44.4′ N, 69°19.2′ W; *d*, Acadia National Park, 44°17.7′ N, 68°10.7′ W. Gulf Stream: *e*, 25°58.5′ N, 79°42.5′ W; *f*, 30°10.4′ N, 79°33.8′ W. Sargasso Sea: *g*, 27°21.4′ N, 73°12.6′ W; *h*, 27°20.9′ N, 73°14.9′ W; *i*, 27°21.4′ N, 73°16.4′ W. Biscayne Bay, Florida: *j*, 25°41.0′ N, 80°08.1′ W.

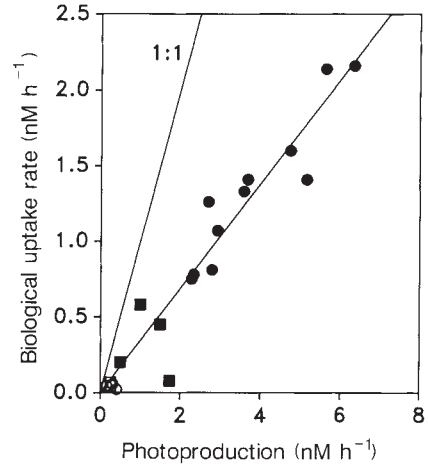


FIG. 2 Biological uptake rate of pyruvate plotted against its midday rate of photochemical production. The line drawn through the data was determined from linear regression analysis. The 1:1 line (slope = 1) depicts the case where photochemical production and biological uptake rates are equal. Different symbols represent different sampling locations: (○) Sargasso Sea, (□) Gulf Stream, (■) Gulf of Maine and (●) Biscayne Bay, Florida.

strongly correlated to its rates of photochemical production: $r = 0.964$, $n = 20$, slope $\pm 1\sigma = 0.35 \pm 0.03$. Based on the inverse of the regression slope, the midday rates of photochemical production are larger than corresponding rates of biological uptake by a factor of approximately three. We expect this relationship to vary both vertically in the water column and as a function of the solar light flux. Imbalances in rates of uptake and photo-production should give rise to diurnal variations in the concentration of pyruvate in surface waters, when other physical factors (such as vertical mixing) are relatively unimportant. We have observed diurnal variations at several coastal locations (D.J.K. *et al.*, manuscript in preparation) and these variations can be simulated readily in the laboratory[4] using filtered and unfiltered sea water. In one field study[7], maximum daytime concentrations were ~4.0 nM whereas night-time concentrations were as low as 0.5 nM. We have observed similar diurnal variations for other photochemically generated biological substrates in sea water, including glyoxylate and formaldehyde; we will discuss the details of these studies elsewhere. These results support the direct relation between the photochemical production and biological uptake of pyruvate shown in Fig. 2. Another important feature depicted in Fig. 2 is that both biological uptake and photochemical production rates are generally higher at coastal locations (Biscayne Bay in Florida and the Gulf of Maine) than at oligotrophic sites in the Gulf Stream and the Sargasso Sea.

Our results clearly show that there is strong correlation between rates of uptake and photochemical production of pyruvate in surface waters. However, these processes may be only incidentally related and may be a consequence of variations in other parameters such as the concentration of dissolved organic matter (DOM). On the other hand, it is quite reasonable to expect that bacteria have adapted to varying concentrations of photochemically produced substrates in the photic zone, such as we observe for pyruvate. Therefore, we expect a direct coupling of uptake rates to production rates, as has been shown for the biological turnover of amino acids using an isotope dilution technique.[9]

We performed ultrafiltration experiments to obtain information regarding the photochemical source(s) of pyruvate in sea water. Filtered sea water collected from the Sargasso Sea or Biscayne Bay was fractioned through a YC05 Amicon ultrafiltration membrane with a nominal molecular-weight cutoff of 500 (ref. 7). Filtered (0.22 μm) Sargasso sea water and Biscayne Bay water showed photochemical production of pyruvate at rates of 1.1 and 5.3 nM W^{-1} hr^{-1} m^{-2} respectively (Fig. 3). By contrast,

we observed no photochemical production of pyruvate for either sample in the fraction of DOM having a nominal molecular weight <500. In fact, we observed a slight loss, as expected, because pyruvate (and other α-keto acids) undergo photochemical decarboxylation in sea water[7] (D.J.K. *et al.*, manuscript in preparation). Although high-molecular-weight DOM seems to be the principal photochemical source of pyruvate, it is not possible, on the basis of ultrafiltration alone, to discern whether the precursors consist of biologically refractory or biologically labile material. There is some evidence, however, that suggests that the precursors are in fact biologically refractory. Sargasso sea water used here was collected from 3,000 m; previous investigators have shown that DOM from this depth is biologically refractory in nature[10-13]. The refractory nature of this DOM is also indicated by its mean residence time (based on ^{14}C measurements) of >6,000 yr (ref. 2).

A critical assumption of our hypothesis is that there exists a large pool of DOM in the oceans that is biologically refractory. There are many lines of evidence supporting this idea (for review see ref. 1); however, these studies are not without uncertainties. Further studies are needed to define more clearly whether photochemical production of biologically labile products originates from a biologically refractory pool of DOM. Does this pathway represent a new input of biologically labile DOM in the photic zone, or does it simply involve transformations of DOM that is already biologically labile?

In the planktonic food web[14], phytoplankton are generally believed to be the major source of DOM[15,16] for bacteria. If the magnitude of photochemical production of biologically labile compounds is large in the sea, particularly with respect to oligotrophic environments, then this photochemical pathway may represent an important non-phytoplankton source of DOM available to bacteria. We have calculated a rough estimate of the photochemical flux of biologically labile DOM, specifically DOC, to heterotrophic bacteria relative to the flux from primary productivity in the Sargasso Sea. Given a total daytime (8 h of sunlight) photochemical production of pyruvate of ~4 nM per day in the upper 10 m, the corresponding DOC input is 1.6 mg C m^{-2} per day. This input is ~1% of the primary productivity[17] (typical value 200 mg C m^{-2} per day) in the Sargasso Sea. Assuming that pyruvate is primarily taken up by bacteria and that 10-50% of primary productivity passes through heterotrophic bacteria on a daily basis[18,19], then the photochemical flux of pyruvate is 2-4% of the total carbon flux to bacteria. Even though this percentage is small, it is based on the photochemical production only of pyruvate. It is reasonable to expect that a wide variety of other easily utilized biological substrates of low molecular weight are formed by the photochemical breakdown of biologically refractory DOM (acetate, acetaldehyde, formate, formaldehyde, glyoxylate, carbon monoxide, methanol, for example); several of these have already been identified[4,20,21]. When the production rates of these compounds are summed, photo-oxidation of biologically refractory DOM may well be a significant source of DOM to bacteria, especially if bacterial growth in oligotrophic environments is carbon-limited in some instances[22]. In addition to the production of common biological substrates, photochemical alteration of the DOM residue may also render it more susceptible to microbiological attack, as has been shown for fresh-water DOM[23]. Depending on the products formed and the relative uptake of these products by different bacteria, these carbon fluxes may be a determinant in the structure of the microheterotrophic community in the upper ocean.

We have shown that photochemical fragmentation of biologically refractory DOM (such as humic substances) results in the formation of biologically labile compounds (such as pyruvate) and that biological uptake of these photochemical products is strongly coupled to their rate of photochemical production in surface sea water. Rough calculations indicate that this overall process is important in the removal of oceanic DOC, and probably also accounts for the rapid degradation of riverine organic carbon inputs into the sea. These calculations, as well as further evidence supporting this latter hypothesis, will be presented elsewhere (K.M. *et al.*, manuscript in preparation). □

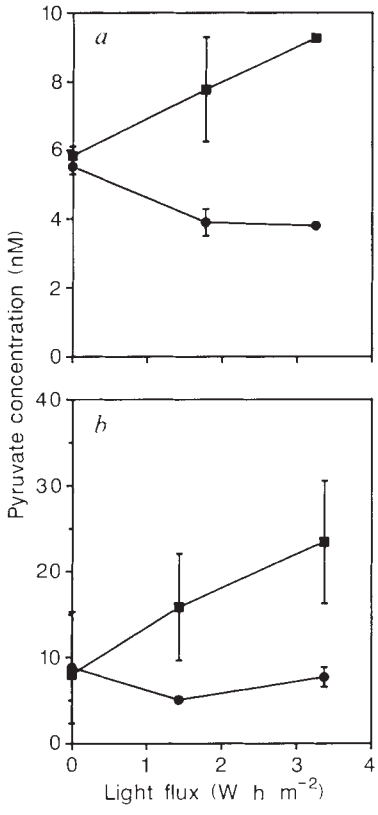


FIG. 3 Plot of the concentration of pyruvate as a function of the near-ultraviolet light flux (<385 nm) in sea water that is 0.22-μm filtered (■) or fractioned by ultrafiltration using a membrane with molecular-weight cutoff of 500 (●). It is assumed that the difference in production rates between these two fractions is equal to the rate of photochemical production in the fraction greater than a nominal molecular weight of 500. *a*, Sargasso sea water collected at 3,000 m, and *b*, Biscayne Bay sea water. The vertical bar represents the range of the data ($n=2$); no vertical bar was drawn when the range of the data was smaller than the diameter of the symbol. Note the difference in the scale of the vertical axis between *a* and *b*.

Received 23 June; accepted 11 September 1989.

1. Mopper, K. & Degens, E. T. in *The Global Carbon Cycle* (eds Bolin, B., Degens, E.T., Kempe, S. & Ketner, P.) 293-316 (Wiley, New York, 1979).
2. Williams, P. M. & Druffel, E. R. M. *Nature* **330**, 246-248 (1987).
3. Williams, P. M. & Druffel, E. R. M. *Oceanography* **1**, 14-17 (1988).
4. Kieber, D. J. & Mopper, K. *Mar. Chem.* **21**, 135-149 (1987).
5. Hobbie, J. E. & Crawford, C. C. *Limnol. Oceanogr.* **14**, 528-532 (1969).
6. Bailey, C. A., Neihof, R. A. & Tabor, P. S. *Appl. Environ. Microbiol.* **46**, 44-49 (1983).
7. Kieber, D. J. thesis, Univ. Miami (1988).
8. Kendall, M. G. & Stuart, A. *The Advanced Theory of Statistics* Vol. 2 (Hafner, New York, 1961).
9. Fuhrman, J. A. *Mar. Ecol. Prog. Ser.* **37**, 45-52 (1987).
10. Menzel, D. W. & Ryther, J. H. *Deep Sea Res.* **15**, 327-337 (1968).
11. Barber, R. T. *Nature* **220**, 274-275 (1968).
12. Craig, H. *J. geophys. Res.* **76**, 5078-5086 (1971).
13. Williams, P. M. & Carlucci, A. F. *Nature* **262**, 810-811 (1976).
14. Williams, P. J. LeB. *Keiler Meeresforsch. Sonderh.* **5**, 1-28 (1981).
15. Sharp, J. H. in *Heterotrophic Activity in the Sea* (eds Hobbie, J. E. & Williams, P. J. LeB.) 101-120 (Plenum, New York, 1984).
16. Azam, F. & Cho, B. C. in *Ecology of Microbial Communities* (eds Fletcher, M., Gray, T. R. G. & Jones, J. G.) 261-281 (Cambridge University Press, Boston, 1987).
17. Menzel, D. W. & Ryther, J. H. *Deep Sea Res.* **6**, 351-367 (1960).
18. Ducklow, H. W. & Hill, S. M. *Limnol. Oceanogr.* **30**, 239-259 (1985).
19. Ducklow, H. W., Purdie, D. A., Williams, P. J. LeB. & Davies, J. M. *Science* **232**, 865-867 (1986).
20. Mopper, K. & Stahovec, W. L. *Mar. Chem.* **19**, 305-321 (1986).
21. Conrad, R. and Seiler, W. *FEMS Microbiol. Lett.* **9**, 61-64 (1980).
22. Fuhrman, J. A., Sleeter, T. & Carlson, C. *Eos* **68**, 1729 (1987).
23. Geller, A. *Limnol. Oceanogr.* **31**, 755-764 (1986).

ACKNOWLEDGEMENTS. We thank R. Cuhel, R. Jones, E. Saltzman, D. Savoie, B. Taylor and R. Zika for discussions and R. Sikorski for technical assistance. We also thank the captain and crew of the RV *Columbus Iselin*. We acknowledge the NSF and the Office of Naval Research for providing shiptime and financial support.

Early incorporation of polysulphides in sedimentary organic matter

M. E. L. Kohnen*, J. S. Sinninghe Damsté*, H. L. ten Haven† & J. W. de Leeuw*

* Organic Geochemistry Unit, Faculty of Chemical Technology and Materials Science, Delft University of Technology, de Vries van Heystplantsoen 2, NL-2628 RZ Delft, The Netherlands
† Institute of Petroleum and Organic Geochemistry (ICH-5), KFA Jülich, PO Box 1913, D-5170 Jülich, FRG

THE increase in sulphur content of organic matter with depth in Recent sediments has been attributed to incorporation of either H_2S (ref. 1) or polysulphides[2,3] or to a combination of both[4,5]. Indeed, the widespread occurrence of organic sulphur compounds[6–12] with carbon skeletons that bear an unambiguous link with natural precursors indicates that organic matter may act as a sink for inorganic sulphur species with an efficacy second only to reactive iron minerals[13]. Laboratory and field observations[14–18] indicate that all compounds identified so far are consistent with incorporation of H_2S; molecular evidence for incorporation of polysulphides has previously been lacking. Here we report the identification of homologous series of cyclic disulphides with a linear carbon skeleton and of a cyclic di- and trisulphide with a C_{20} isoprenoid carbon skeleton in sediments of Quaternary to Pliocene age. Although incorporation of H_2S can still explain the presence of cyclic disulphides, the cyclic trisulphide implies incorporation of inorganic polysulphides in sedimentary organic matter at the earliest stages of diagenesis.

Several organic-carbon-rich sediment samples of Quaternary to Pliocene age from the Peruvian upwelling area[19] and a bituminous shale of Messinian age (upper Miocene) from the Vena del Gesso basin (located in the northern Apennines[20]) were extracted. Gas chromatographic analyses with dual flame ionization detection and flame photometric detection of the aromatic hydrocarbon fraction isolated from the extracts[11,19] revealed the presence of compounds containing more than one sulphur atom.

In the Vena del Gesso sample, two homologous series of compounds with two sulphur atoms are most abundant, whereas in several of the Peruvian samples a compound with three sulphur atoms dominates the aromatic hydrocarbon fractions together with the polycyclic aromatic hydrocarbon perylene. The major homologous series occur as doublet peaks (Fig. 1*a*) with identical mass spectra, all characterized by an intense fragment ion at $M^{+\cdot}-33$ (–SH, based on exact mass measurements), a minor ion at $m/z=133$ ($C_5H_9S_2$) and several secondary-fragment ions (for example, $m/z=101$, 143 and 185) which are also observed in mass spectra of 5-alkyl-2-methylthiolanes[8,10,21]. Loss of SH from the molecular ion has been reported for dithianes[22,23]. The major products of Raney nickel desulphurization[10] of the sulphide fraction (isolated by argentatious thin-layer chromatography), followed by hydrogenation, were *n*-alkanes. These characteristics indicate that a series of stereoisomers of 3-*n*-alkyl-6-methyl-1,2-dithianes (C_{15}–C_{24}) occurs in this sample. Synthesis of the major members (*cis* and *trans* 3-methyl-6-tridecyl-1,2-dithiane, C_{18}; compound I), for which details are described elsewhere[24], and subsequent co-injections on capillary columns of different polarity confirmed this assignment.

Mass chromatography of the molecular ions of the cyclic disulphides ($274+14n$) revealed the presence of C_{20} compounds (compound II, Fig. 1*b*), eluting much earlier than the C_{20} dithianes with linear carbon skeletons. The only notable difference in their mass spectra is the shift of fragment ion $m/z=143$ to $m/z=157$, a characteristic ion of 3-methyl-2-(3,7,11-trimethyldodecyl)thiolane[10]. Raney nickel desulphurization also gave a substantial amount of phytane. Hence these compounds are believed to be *cis* and *trans* 4-methyl-3-(3,7,11-trimethyldodecyl)-1,2-dithianes (compound II). A third isoprenoid cyclic disulphide is tentatively identified as 4-(4,8,12-trimethyltridecyl)-1,2-dithiane (compound III) based on its gas-chromatographic retention time and mass spectrum. These

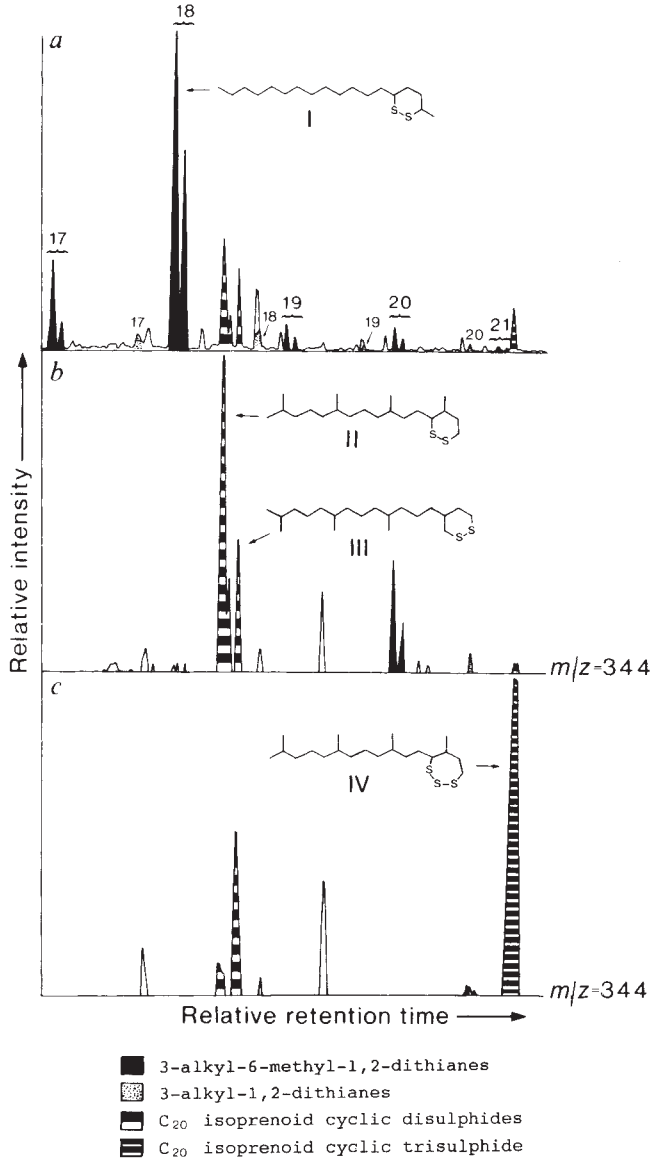


FIG. 1 *a*, Partial reconstructed ion chromatogram of the aromatic hydrocarbon fraction of bituminous shale from the Vena del Gesso basin. *b*, Partial-mass chromatograms of $m/z=344$ of the aromatic hydrocarbon fractions of a bituminous shale from the Vena del Gesso basin. *c*, Mottled diatomaceous ooze from the Peruvian upwelling area (ODP 112-684C-4H-7, 30–37 cm; 36.15 m sub-bottom depth). The carbon numbers of the homologous series of *cis* and *trans* 3-*n*-alkyl-6-methyl-1,2-dithianes and 3-*n*-alkyl-1,2-dithianes are indicated in *a*. Minor compounds, not indicated in *a*, are mainly homologous series of 2-*n*-alkylthiophenes and 2-*n*-alkyl-5-methylthiophenes. Compounds I to IV are discussed in the text. Instrumental conditions were: a Carlo Erba Fractovap 4160 gas chromatograph equipped with a Hewlett-Packard fused-silica capillary column (length, 50 m; inner diameter, 0.31 mm) coated with Ultra 2 (film thickness, 0.52 μm). Temperature increased from 110 °C to 320 °C at a rate of 3 °C min^{-1}. Helium was used as the carrier gas. The gas chromatograph was connected to a VG-7070E mass spectrometer operating at 70 eV. Exact mass measurements were performed by peak matching using selected masses from perfluorkerosine as the reference and a resolution of 5,000 with a VG-70S mass spectrometer operating at an ionization voltage of 70 eV.

isoprenoid cyclic disulphides were also present in most of the Peruvian samples, although the linear 1,2-dithianes were virtually absent (Fig. 1*c*). However, the most important extractable organic sulphur compound (compound IV, which is sometimes the most abundant compound of the aromatic hydrocarbon fraction) gives a mass spectrum with a molecular ion at $m/z = 376$ ($C_{20}H_{40}S_3$; based on exact mass measurements), an ion at $m/z = 344$ ($M^{+} - 32$), and a fragmentation pattern similar to that observed for the C_{20} isoprenoid 1,2-dithianes ($m/z = 101$, 157, 255 and 311). Loss of sulphur from a parent ion ($M^{+} - 32$) has been reported for molecules with two or more adjacent sulphur atoms[21,25-27]. The product of Raney nickel desulphurization of the aromatic hydrocarbon fraction is almost solely phytane. This compound was therefore identified as 5-methyl-4-(3,7,11-trimethyldodecyl)-1,2,3-trithiepane (compound IV). This cyclic trisulphide also occurs as a minor compound in the bituminous shale of the Vena del Gesso basin (Fig. 1). We note that the sporadic occurrence of compound IV in the Peruvian sediments excludes its formation after sample preparation.

The structures of these novel organic sulphur compounds cannot be related to known naturally occurring di- and trisulphides[28]. By analogy with the hypothesis put forward to explain the origin of other organic sulphur compounds[6-11], the incorporation of abiotic sulphur into specific functionalized lipids can explain their occurrence. For example, phytadienes (derived from ubiquitous phytol) can act as suitable precursors for the newly identified C_{20} isoprenoid cyclic di- and trisulphides. In contrast to previously suggested incorporation models[11], addition of H_2S cannot be invoked to explain the occurrence of a C_{20} isoprenoid cyclic trisulphide. The cyclic disulphides may still originate from reactions of H_2S through intermediate dithiols which, after oxidation, may yield 1,2-dithianes. Chemically and/or microbially generated inorganic polysulphides[29,30] are suitable sulphur species for the formation of cyclic trisulphides, although also microbially produced polysulphides occur the most prominent ones occurring in natural aqueous solutions[29] are HS_4^-, HS_5^-, S_4^{2-} and S_5^{2-}.

It is known[28,31] that organic polysulphides are thermally unstable and therefore we suggest that if tetra- and pentasulphides are formed they will easily degrade to tri- and disulphides during prograde diagenesis. This also implies that the latter compounds may be seen as potential precursors for sulphides (thiolanes). We note that the younger and less deeply buried Peruvian sediment samples contain the trisulphide in high abundance in comparison to disulphides, a situation which is reversed in the older and once more deeply buried bituminous shale from the Vena del Gesso basin (Fig. 1).

Di- and trisulphides may prove to be useful tools for the reconstruction of palaeoenvironments because the occurrence of inorganic polysulphides is known to be sensitive to variations of the pH and redox potential Eh[29]. □

Received 13 July; accepted 7 September 1989.

1. Nissenbaum, A. & Kaplan, I. R. *Limnol. Oceanogr.* **17,** 570-582 (1972).
2. Aizenshtat, Z., Stoler, A., Cohen, Y. & Nielsen, H. in *Adv. org. Geochem. 1981* (eds Bjorøy, M. *et al.*) 279-288 (Wiley, Chichester, 1983).
3. LaLonde, R. T., Ferrara, L. M. & Hayes, M. P. *Org. Geochem.* **11,** 563-571 (1987).
4. Francois, R. *Geochim. cosmochim. Acta* **51,** 17-27 (1987).
5. Nriagu, J. O. & Soon, Y. K. *Geochim. cosmochim. Acta* **49,** 823-834 (1985).
6. Valisolalao, J. Perakis, N., Chappe, B. & Albrecht, P. *Tetrahedron Lett.* **25,** 1183-1186 (1984).
7. Brassell, S. C., Lewis, C. A., de Leeuw, J. W., de Lange, F. & Sinninghe Damsté, J. S. *Nature* **320,** 160-162 (1986).
8. Sinninghe Damsté, J. S., ten Haven, H. L., de Leeuw, J. W. & Schenck, P. A. in *Adv. org. Geochem. 1985* (eds Leythaeuser, D. & Rullkötter, J.), 791-805 (Pergamon, Oxford, 1986).
9. Sinninghe Damsté, J. S., Kock-van Dalen, A. C., de Leeuw, J. W. & Schenck, P. A. *Tetrahedron Lett.* **28,** 957-960 (1987).
10. Sinninghe Damsté, J. S. *et al. Geochim. cosmochim. Acta* **51,** 2369-2391 (1987).
11. Sinninghe Damsté, J. S., Rijpstra, W. I. C., Kock-van Dalen, A. C., de Leeuw, J. W. & Schenck, P. A. *Geochim. cosmochim. Acta* **53,** 1343-1355 (1989).
12. ten Haven, H. L., Rullkötter, J., Sinninghe Damsté, J. S. & de Leeuw, J. W. in *Geochemistry of Sulfur in Fossil Fuels* (eds Orr, W. L. & White, C. M.) (American Chemical Society, Washington, DC, in the press).
13. Berner, R. A. *Geochim. cosmochim. Acta* **48,** 605-615 (1985).
14. Casagrande, D. J. & Ng, L. *Nature* **282,** 598-599 (1979).
15. Casagrande, D. J. *et al. Nature* **282,** 599-600 (1979).
16. Mango, F. D. *Geochim. cosmochim. Acta* **47,** 1433-1441 (1983).
17. Vairavamurthy, A. & Mopper, K. *Nature* **329,** 623-625 (1987).
18. Moers, M. E. C., de Leeuw, J. W., Cox, H. C. & Schenck, P. A. in *Adv. org. Geochem. 1987* (eds Mattavelli, L. & Novelli, L.) 1087-1091 (Pergamon, Oxford, 1988).
19. ten Haven, H. L. Littke, R., Rullkötter, J., Stein, R. & Welte, D. H. in *Proc. Ocean Drilling Prog., Sci. Results, 112* (eds Suess, E. *et al.*) (Ocean Drilling Program, College Station, Texas, in the press).
20. Vai, G. B. & Ricci Lucchi, F. R. *Sedimentology* **24,** 211-244 (1977).
21. Schmid, J. C., Connan, J. & Albrecht, P. *Nature* **329,** 54-56 (1987).
22. Bowie, J. H. *et al. J. chem. Soc.* B 946-951 (1966).
23. Bowie, J. H. & White, P. Y. *Org. Mass Spectrom.* **6,** 317-324 (1972).
24. Kohnen, M. E. L., Sinninghe Damsté, J., ten Haven, H. L., Kock-van Dalen, A. C. & de Leeuw, J. W. *Geochim. cosmochim. Acta* (submitted).
25. Moriaty, R. M., Ishibe, N., Kayser, M., Ramey, K. C. & Gisler, H. J. Jr *Tetrahedron Lett.* **55,** 4883-4888 (1969).
26. Schwab, A. W., Rohwedder, W. K. & Ard, J. S. *Phosphorus Sulf.* **7,** 315-320 (1979).
27. Singh, P. K., Field, L. & Sweetman, B. J. *Phosphorus Sulf.* **39,** 61-71 (1988).
28. Field, L. in *Organic Chemistry of Sulfur* (ed. Oae, S.) 303-382 (Plenum, New York, 1977).
29. Boulegue, J., Lord, C. J. & Church, T. M. *Geochim. cosmochim. Acta* **46,** 453-464 (1982).
30. Luther, G. W., Church, T. M., Giblin, A. E. & Howarth, R. W. in *Organic Marine Geochemistry* (ed. Sohn, M.) 340-355 (American Chemical Society, Washington, DC, 1986).
31. Pickering, T. L. & Tobolsky, A. V. in *Sulfur in Organic and Inorganic Chemistry* Vol. 3 (ed. Senning, A.) 19-38 (Marcel Dekker, New York, 1972).

ACKNOWLEDGEMENTS. We thank J. Ebbing for the Vena del Gesso sample and the international Ocean Drilling Program for the samples from ODP Leg 112. This work was partly supported by the Netherlands Foundation for Earth Science Research with financial aid from the Netherlands Organisation for Scientific Research and the Deutsche Forschungsgemeinschaft.

Phytoplankton production pulses and episodic settlement of a temperate marine fish

Ronald E. Thresher, Graham P. Harris, John S. Gunn & Lesley A. Clementson

CSIRO Marine Laboratory, GPO Box 1538, Hobart, Tasmania 7001, Australia

THE settlement of post-larval marine invertebrates[1,2] and fishes[3-5] fluctuates massively both within and between years; this variability significantly affects the structure of adult populations[5,6]. The causes of these fluctuations are not well known, but they are typically attributed to variation in some element of the planktonic environment that determines the distribution and rates of survival of the larvae[3,7-9]. Here we report that over a three-year period, pulses of settlement of a species of temperate reef fish were invariably preceded by brief, irregularly occurring peaks of phytoplankton production. The lag time was consistent with a 'critical period' hypothesis[10], in which settlement rates are determined by irregular variation in the availability of food for newborn larvae. Although they are difficult to detect, transient but intense pulses of production may be common in the oceanic water column[11] and could underlie the extreme temporal variability of settlement characteristic of many marine organisms.

The viviparous rocky reef fish *Heteroclinus* sp. ("Scott's Weedfish" in ref. 12), occurs in large numbers in tide pools and shallow reef areas along the coast of southern Australia. During a parturition season that lasts from October to late December, females produce clutches of up to 2,700 larvae asynchronously about every 17 days (J.S.G. and R.E.T., unpublished manuscript). At birth, the larvae vary from 7.1-9.4 mm Standard Length, are active swimmers and are well developed, with functional jaws and eyes, no evident yolk sac and an open gut. The larvae of this and other clinid species are found predominantly inshore, in coastal bays and inlets. The duration of the planktonic larval stage varies seasonally from 3 to 6 weeks, after which the fish settle as lightly pigmented juveniles, 14-16 mm Standard Length, in tide pools and shallow rocky habitats.

We measured seasonal variability in settlement by *Heteroclinus* sp. for three years (1985/86 to 1987/88) by sampling every two weeks at sites around Storm Bay, a large semi-autonomous[13] embayment in southeastern Tasmania (43° 20′S, 147° 30′ E). The date on which each juvenile settled was determined by examination of otolith microstructure[14]. Over the same period we measured phytoplankton abundance weekly at a station near

the centre of Storm Bay. Phytoplankton abundance was quantified by chlorophyll levels, using the hot-methanol extraction method[13].

The number of newly settled *Heteroclinus* sp. collected per sample varied widely and irregularly each spawning season. The patterns of temporal variation in settlement differed markedly between years. One large prolonged peak dominated settlement in 1985/86; two small peaks, separated by several weeks of almost no settlement, were evident in 1986/87; in 1987/88, there was little settlement early in the year, but thereafter settlement was continuous, with two peak periods separated by about 5 weeks (Fig. 1, top).

The temporal pattern of settlement each year matched that of chlorophyll levels in Storm Bay several weeks earlier (Fig. 1, bottom). Over the three-year period, all the main peaks in settlement could be matched to a corresponding peak in chlorophyll levels 7-9 weeks earlier (modally 7 weeks); conversely, there were no peaks in chlorophyll levels that were not followed by a pulse of settlement. A Spearman rank correlation between numbers of settling fishes and preceding chlorophyll levels is significant ($P<0.05$) in each of the three years, at identical lag times (the correlations are significant at lags of 6, 7 and 8 weeks in 1985/86, 7 and 8 weeks in 1986/87, and 7 and 8 weeks in 1987/88). The correlation between chlorophyll levels and settlement, pooled by calendar week, accounts for 56%, 37% and 23% of the variance in settlement for the 1985/86, 1986/87 and 1987/88 years, respectively.

By contrast, chlorophyll levels do not correlate significantly with subsequent settlement when years are pooled (maximum Spearman rank correlation coefficient $r_s=0.23$, $n=40$, lag of 8 weeks), reflecting a poor relationship between year-class strength (total number of newly settled fishes per annum) and mean annual chlorophyll levels. Nor is there any apparemt relationship among years between year-class strength and the minimum, maximum, variability (coefficient of variation) or sum of chlorophyll levels each year. Across years, settlement peaks coincide (after lagging) with chlorophyl levels greater than $\sim 2\mu g^{-1}$, but the magnitude of settlement often shows no relation to the magnitude of the corresponding peak in chlorophyll.

The lag between chlorophyll level and settlement is consistent with the hypothesis that variations in food availability for newborn larvae determine settlement rates. Gut contents of *Heteroclinus* larvae less than 5-days-old consist almost entirely of small copepods and similar sized microzooplankton. The abundance of these microzooplankton in the weekly samples correlates strongly with that of phytoplankton, peaking 1-4 weeks (highest cross-correlation at a lag of 2 weeks) after each phytoplankton peak (G.P.H., L.A.C. and F. B. Griffiths, manuscript in preparation). These larvae spend, on average, 4 weeks in the plankton ($\bar{x}=28.8$ days, $n=511$, years pooled), varying seasonally from a mean of ~6 weeks early in the spawning season to 3.5 weeks in mid-summer. The sum of the chlorophyll-microzooplankton lag and the duration of the fish's larval stage, therefore, ranges from ~4.5-9 weeks, and averages 6-7 weeks. This compares with a correlation between chlorophyll level and settlement at lags that range from 6-8 weeks and average 7 weeks.

It is not possible to falsify the hypothesis that the temporal patterns of production and settlement are independent responses to a third, forcing function. Nonetheless, we consider this possibility unlikely. The timing of chlorophyll peaks in Storm Bay is determined by episodic bouts of strong winds, which turn over the water column and result in nutrient enrichment and, about three weeks later, a phytoplankton bloom (G.P.H., L.A.C. and F. B. Griffiths, manuscript in preparation). Hence the physical event responsible for chlorophyll peaks occurs, on average, 10 weeks before the settlement peaks and 5 weeks before the birth of the larvae that constitute those peaks. If wind events directly determine settlement variation, then they would have to operate by affecting pregnant females, producing short-term variability in brood size or larval competence. Brood sizes, however, are set before the onset of the parturition season, and neonate size, at least, varies only slightly seasonally and does not correlate with settlement variability (J.S.G. and R.E.T., unpublished manuscript). Alternative hypotheses based on, for example, the effects of wind on patterns of larval advection are possible, but difficult to reconcile with the lag times observed. Moreover, water-mass characteristics indicate that exchange in and out of Storm Bay is relatively slow[13] and that the large scale episodic flushing that would be required to cause marked variability in settlement is unlikely.

Observations of the larvae in aquaria also support the hypothesis that short-term variations in food availability directly mediate the link between water-column production and settlement rate. As yet, we have been unable to get newborn *Heteroclinus* larvae to feed; all but a few of the several thousand larvae that we have produced in the laboratory died, apparently of starvation, within 12 h after birth. The short period to starvation is consistent with the lack of yolk sacs of these larvae at birth[15]. Older larvae, however, collected in the field and held under the same conditions, fed readily and metamorphosed as expected.

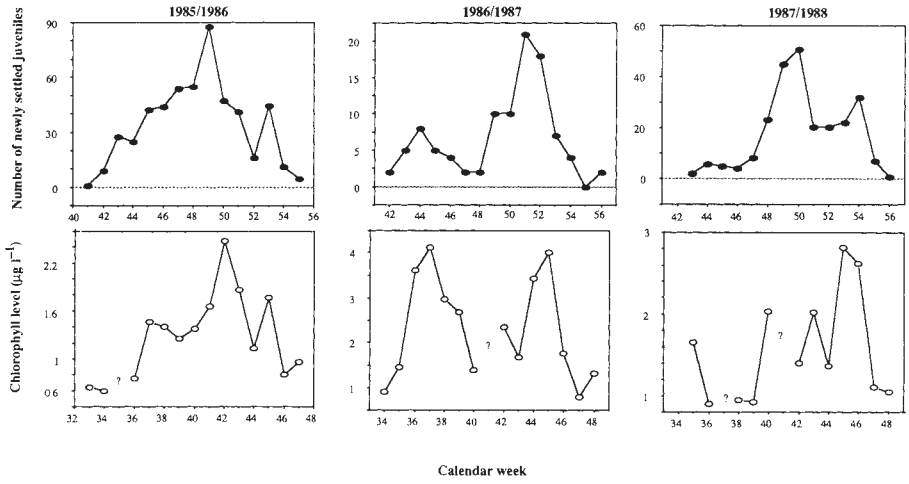


FIG. 1 Total numbers of newly settled *Heteroclinus* sp. recruiting to three sampling sites in Storm Bay, pooled by calendar week, in the 1985/86, 1986/87 and 1987/88 reproductive seasons. Note the different scales on the ordinates. Bottom, temporal variation in chlorophyll levels 8 weeks before the beginning of the reproductive season each year of *Heteroclinus* sp. measured at a weekly sampling station near the centre of Storm Bay. Question marks represent missing data. Note that the first main peak in settlement in 1987/88 occurred 8 weeks after a missing datum for chlorophyll levels. Although we cannot be certain of a production peak that week, chlorophyll levels rose abruptly the week before the missing sample (as did settlement 8 weeks later), and the missing sample was preceded by a short duration nutrient pulse (see ref. 13), both of which are consistent with a peak in chlorophyll during the missing period.

If *Heteroclinus* larvae must eat very soon after birth, as our experience and the absence of yolk reserves indicates, then it is plausible that short-term variations in microzooplankton abundance could markedly affect the numbers of larvae that survive the first 24 h after birth, along the lines of the critical period hypothesis previously suggested to account for inter-annual variations in recruitment success[10].

Short-term transients of nutrient enrichment and subsequent phytoplankton production, on time scales of hours to days, can result from a range of oceanographic features, including meso-scale eddy formation[11], storm surges[16], current meanders[17] and tidally induced internal density waves (solitons)[18]. Although difficult to detect because of the temporal and spatial limitations imposed by ship-based sampling[11,19], these transients have recently been suggested to have importznt consequences for biogeochemical cycling and productivity in the ocean[19,20]. Our data indicate that these transients also underlie the extreme temporal patchiness of settlement by marine organisms and may have profound effects on the population ecology of such organisms. □

Received 16 May; accepted 6 September 1989.

1. Connel, J. H. *Ecol. Monogr.* **31,** 61 (1961).
2. Caffey, H. M. *Ecol Monogr.* **55,** 313 (1985).
3. Cushing, D. H. *Marine Ecology and Fisheries* (Cambridge University Press, 1975).
4. Sale, P. F., Doherty, P. J., Eckert, G. J., Douglas, W. A. & Ferrel, D. J. *Oecologia Berlin* **64,** 191 (1984).
5. Doherty, P. J. & Williams, D. M. *Oceanogr. mar. Biol. A. Rev.* **26,** 487-551 (1988).
6. Keough, M. J. *J. exp. mar. Biol. Ecol.* **66,** 213 (1983).
7. Lasker, R. *Rapp. P.-v. Réun. Consperm. int. Explor. Met* **178,** 375 (1981).
8. Parrish, R. H., Nelson, C. S. & Bakun, A. *Biol. Oceanogr.* **1,** 175 (1981).
9. Houde, E. D. *Am. Fish. Soc. Symp. No. 2.* **17,** (1987).
10. Hjort, J. J. *Rapp. P.-v. Réun Cons. perm. Int. Explor. Met* **1,** 1 (1926).
11. Lohrenz, S. E., Arnone, R. A., Wiesenberg, D. A. & DePalma, I. P. *Nature* **335,** 245-247 (1988).
12. Last, P. R., Scott, E. O. G. & Talbot, F. H. *Fishes of Tasmania* (Tasmanian Fisheries Development Authority, Hobart, Tasmania, 1983).
13. Clementson, L. A., Harris, G. P., Griffiths, F. B. & Rimmer, D. W. *Aust. J. mar. Freshwat. Res.* **40,** 25 (1989).
14. Campana, S. E. & Neilson, J. D. *Can. J. Fish. aquat. Sci.* **42,** 1014 (1985).
15. McGurk, M. D. *Mar. Biol.* **84,** 13 (1984).
16. Checkley, D. M., Raman, S., Maillet, G. L. & Mason, K. M. *Nature* **335,** 346-348 (1988).
17. Andrews, J. C. & Gentien, P. *Mar. Ecol. Prog. Ser.* **8,** 257-269 (1982).
18. Holligen, P. M., Pingree, R. D. & Mardell, G. T. *Nature* **314,** 348-350 (1985).
19. Jenkins, W. J. *Nature* **331,** 521-523 (1988).
20. Platt, T. *Mar. Ecol. Prog. Ser.* **52,** 77-88 (1989).

ACKNOWLEDGEMENTS. We thank C. Bobbi, B. Bruce, D. Furlani, B. F. Griffiths and D. Rimmer for their assistance during the study, A. Gronell and K. Sainsbury for discussions of the results, and D. H. Cushing, F. R. Harden Jones, R. Johannes, V. Mawson and P. Young for commenting on the manuscript.

Cortical magnification factor and the ganglion cell density of the primate retina

Heinz Wässle, Ulrike Grünert, Jürgen Röhrenbeck & Brian B. Boycott*

Max-Planck-Institut für Hirnforschung, Deutschordenstrasse 46, D-6000 Frankfurt 71, FRG
* MRC Cell Biophysics Unit, King's College, 26-29 Drury Lane, London WC2B 5RL, UK

IT has long been contentious whether the large representation of the fovea in the primate visual cortex (V1)[1–8] indicates a selective magnification of this part of the retina[7,9–11], or whether it merely reflects the density of retinal ganglion cells[12–14]. The measurement of the retinal ganglion-cell density is complicated by lateral displacements of cells around the fovea and the presence of displaced amacrine cells in the ganglion cell layer. We have now identified displaced amacrine cells by GABA immunohistochemistry and by retrograde degeneration of ganglion cells. By reconstructing the fovea from serial sections, we were able to compare the densities of cones, cone pedicles and ganglion cells; in this way we found that there are more than three ganglion cells per foveal cone. Between the central and the peripheral retina, the ganglion cell density changes by a factor of 1,000–2,000, which is within the range of estimates of the cortical magnification factor[1–8]. There is therefore no need to postulate a selective magnification of the fovea in the geniculate and/or the visual cortex.

There are three main problems in quantifying the ganglion cell density gradient of a primate retina. First, the population of displaced amacrine cells has to be estimated. In all mammalian retinae it is difficult to distinguish displaced amacrine cells from ganglion cells[15,16]. Second, in the fovea, the spatial offset between cone inner segments and ganglion cells due to the cone fibres (Henle fibres, Fig. 2*e*) makes it difficult to estimate the highest sampling density of ganglion cells[17–19]. Finally, the density of cells in the foveal region changes so rapidly with eccentricity that pooling data from experiments involving different animals, as well as different methods and shrinkage factors, produces inaccuracies. In the study reported here we have attempted to overcome these difficulties.

To differentiate ganglion cells from displaced amacrine cells, ganglion cells were retrogradely degenerated by interrupting their axons, and the surviving neurons were counted (Fig. 1*a*, *b*). The density of these neurons decreased from 3,500 cells mm^{-2} close to the fovea to ~500 cells mm^{-2} in the peripheral retina (Fig. 3). Selected fields of this retina were processed for GABA (γ-aminobutyric acid)-like immunoreactivity. Between 80 and 90% of the surviving neurons were stained. The number of neurons in the ganglion cell layer (GCL), which expressed GABA-like immunoreactivity, was also counted in a normal monkey retina (Fig. 1*d*). Close to the fovea a peak of 3,000 cells mm^{-2} was found, and the number decreased to ~500 cells mm^{-2} in the peripheral retina. These data show that most (80–100%) of the neurons that survive retrograde degeneration are GABA-ergic displaced amacrine cells.

Foveal topography was studied from the distribution of cone inner segments, cone pedicles and ganglion cells within the central 3 mm of one monkey retina. Cone density (Fig. 2*a*) decreased from a peak of 250,000 mm^{-2} in the fovea to ~11,500 mm^{-2} at 3 mm eccentricity. There are very few cone pedicles in the centre of the fovea (Fig. 2*b*). By 200 μm eccentricity, their density increases steeply, flattens out at 500 μm eccentricity and peaks in a plateau at 800 μm eccentricity, corresponding to a density of 32,000 mm^{-2}. Close to the fovea the pedicles tend to be elongated circumferentially and are densely packed in a regular way (Fig. 1*c*). The ganglion cell density curve (Fig. 2*c*) is comparable to the density curve for the cone pedicles; up to 2.5 mm eccentricity, however, there are substantially more ganglion cells than cone pedicles (see Fig. 2*d*). There is a plateau at 800 μm eccentricity corresponding to a maximum ganglion cell density of 60,000 mm^{-2}. These data are not corrected for shrinkage; the linear shrinkage factor was 0.9 and the areal was 0.8.

Figure 2*e* represents the geometry of the fovea. The foveal cone pedicles are displaced from their inner segments by Henle fibres as long as 350 μm[17–19]. Between cone pedicles and ganglion cells there is a further displacement of ~50 μm owing to obliquely running bipolar processes and ganglion cell dendrites[19]. These displacements must be allowed for to determine which ganglion cells are connected to which cones. As illustrated in Fig.2*e*, it is possible to get the ratio of ganglion cells to cones of the fovea from a cumulative count. The most central 10,000 cones must be connected to the most central 10,000 cone pedicles, because the Henle fibres have an orderly mapping[19–21]. The radius of the circle that includes these 10,000 cone pedicles

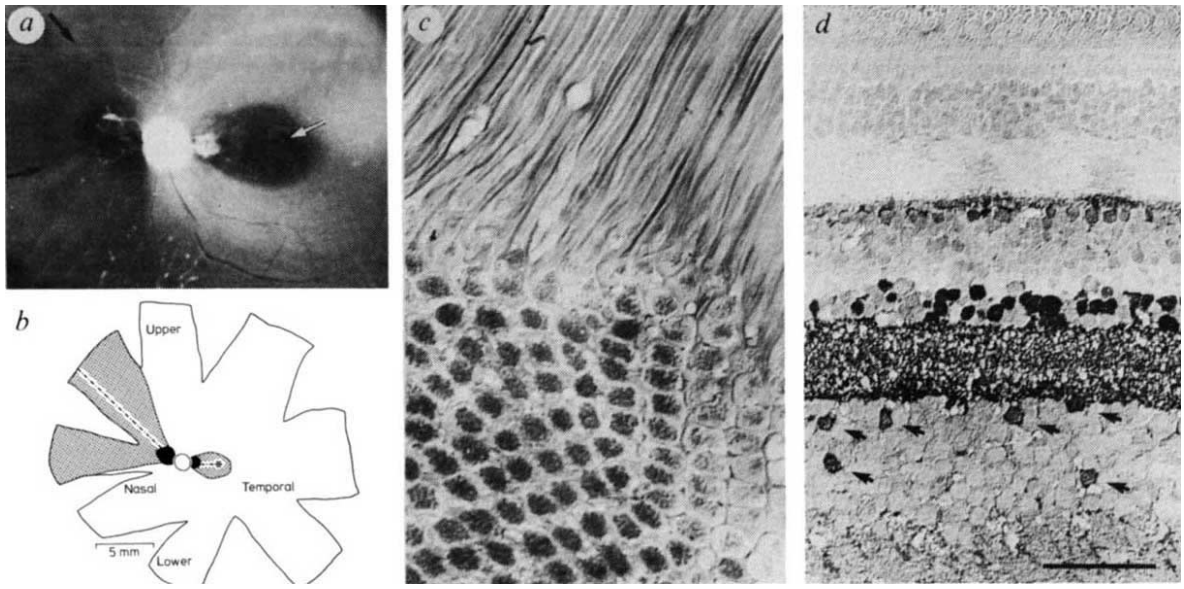


FIG. 1 *a*, Left eye cup of a macaque monkey with retrograde degeneration of ganglion cells produced by two small lesions. They were made in the intact eye using a xenon photocoagulator; the animal was anaesthetized with Ketanest and Rompun. The right eye was left intact, and during the 4-month survival period the animal showed no obvious visual deficiencies. It was killed by an overdose of nembutal and the eyes were enucleated and opened. The white spot in the centre is the optic-nerve head flanked by the two lesions (marked black in *b*). In the temporal retina, an oval black area including the fovea (white arrow) is apparent: the shiny optic-nerve fibre layer has degenerated. The nasal retinal lesion gave a greyish sector of degeneration (black arrows). *b*, Schematic drawing of the lesions (black spots) and retrograde degenerations of ganglion cells (hatched areas). The dashed line gives the transect from fovea to peripheral nasal retina taken to count the surviving neurons (displaced amacrine cells). *c*, Cone pedicles (bottom half) and Henle fibres (top half) in a horizontal section at 500 μm from the fovea. The plane of section through the pedicles and Henle fibres was slightly oblique. The pedicles are regularly arrayed. Close to the fovea they are elongated perpendicularly with respect to the Henle fibres, and up to an eccentricity of 3 mm they are closely packed. *d*, Vertical section through a monkey retina stained for GABA-like immunoreactivity[26]. The section was taken at 1 mm eccentricity, where the ganglion cell layer is 5–6 cell bodies thick; the unstained cell bodies can be seen in this micrograph by using Nomarski optics. Six displaced amacrine cells (arrowed) expressed GABA-like immunoreactivity. Many amacrine cells and horizontal cells stained in the inner nuclear layer (scale bar in *d* represents 3 mm in *a*, 20 μm in *c* and 50 μm in *d*).

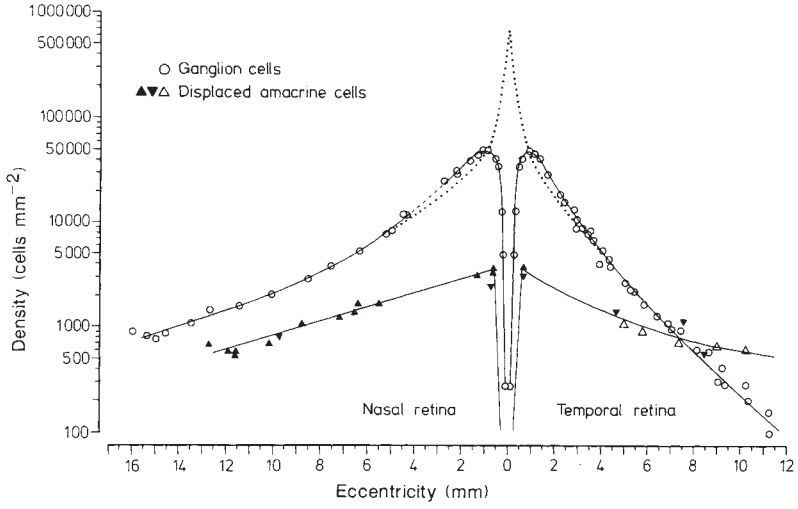


FIG. 3 Densities of ganglion cells (circles) and displaced amacrine cells (triangles) along a horizontal intersect through the macaque monkey retina. The data are corrected for shrinkage. Ganglion cell densities beyond 3-mm eccentricity were measured from Nissl-stained whole-mounts. The data in the temporal retina were measured in three retinae; those in the nasal retina are from two retinae. All neurons in the ganglion cell layer were counted. Then the density of displaced amacrine cells was substracted to give the ganglion cell densities shown. The central 3 mm represents ganglion cell counts taken from sections and corrected for shrinkage. The dotted curve, with a peak corresponding to the foveal centre, shows the sampling density of ganglion cells. By the extension of the cumulative counts (Fig. 2*f*) to successively larger circles, ganglion cell to cone pedicle ratios were calculated at increasing eccentricities. They decreased continuously from 3.34:1 in the fovea to 2:1 at 0.5–1-mm eccentricity and 1:1 at 3–4-mm eccentricity. Hence the sampling density of ganglion cells in the central retina can be derived from the cone densities of Fig. 2*a* by multiplication with the ganglion cell to cone ratios. Filled triangles (tip upwards) give densities of displaced amacrine cells in the lesioned sectors of the retina (Fig. 1*a*, *b*). A rather complete density gradient from the fovea to the nasal retina could be measured. Filled triangles (tip downwards) give densities of displaced amacrine cells, which expressed GABA-like immunoreactivity. Open triangles in the temporal retina represent counts of putative displaced amacrine cells in Nissl stained non-lesioned parts of the retina. Displaced amacrine cells in peripheral monkey retina can be separated from ganglion cells by the criteria used for the cat retina[15,16], except that closer than 5-mm eccentricity the identification of cells becomes uncertain. In three further normal monkey eyes, GABA-immunocytochemistry was performed and the essential findings of the degenerated retina were confirmed: close to the fovea, displaced amacrine cells account for ~5% of all neurons in the ganglion cell layer[32]; in peripheral retina (with the exception of the nasal retina) there are more displaced amacrine cells than ganglion cells.

can be calculated. Adding 50 μm to this radius to allow for the bipolar/ganglion cell displacement gives a circle, which includes the ganglion cells innervated by the most central 10,000 cones. Figure 2*f* shows the curves and the actual integration. In the central 300 μm of the fovea, 9,530 cones were found; in a circle of 500 μm radius around the fovea, there was about the same number of cone pedicles (9,874). The sum of all the ganglion cells in a circle of 550 μm was 33,000, which gives an average of 3.34 ganglion cells per foveal cone.

There should be two midget ganglion cells for every foveal cone, one ON- and one OFF-centre cell. When Pα and Pγ cells (20% of all ganglion cells) are considered, then at least 2.5 ganglion cells per cone must be postulated. We measured 3.34 ganglion cells per foveal cone, which makes it possible that in addition to midget cells there are other Pβ cells, which receive inputs from more than one cone.

A precise reconstruction of the fovea from serial sections was only made in the two retinae of one monkey (Fig. 2). The crucial parameters, however, which are the number of cones in the foveal centre and the ratio of ganglion cells to cone pedicles at their peaks at 750 μm eccentricity, were assessed in three additional retinae. Cone peak density substantially varied (150,000–250,000 mm^{-2}). The number of cones within the central 300-μm (Fig. 2*f*) was less variable (8,536–10,221). Both ganglion cell and cone pedicle densities peaked at 700–800 μm eccentricity, at which their ratio was 1.8 : 1–2 : 1. At smaller eccentricities this ratio was higher. Hence, although there was no reconstruction of the fovea for these three retinae, they seem to have had more than two ganglion cells for every foveal cone. Recent reports from *Macaca mulatta*[18] and from human retina[30] support this result.

Ganglion cells in more peripheral retina were counted from whole mounts. The overall ganglion cell density (corrected for shrinkage) is shown in Fig. 3. In temporal retina, from the maximum at 800 μm eccentricity, the density decreases to a minimum of 100 ganglion cells mm^{-2} at 12 mm eccentricity. The density in the nasal retina is substantially higher beyond ~2 mm eccentricity and does not fall below 800 cells mm^{-2} (ref. 11). The dotted curve in the centre of Fig. 3 gives the sampling density of ganglion cells based on the cumulative cone pedicle and ganglion cell counts (Fig. 2*e*, *f*); the density peaks at 668,250 cells mm^{-2}. A circle with a radius of 200 μm (1°) centred on the fovea is sampled by 38,000 ganglion cells; a circle with a radius of 1 mm (5°) by 220,000 ganglion cells. For comparison, the density gradient of displaced amacrine cells is shown; it changes more gradually from the centre to the periphery. Consequently, in the central retina only 5% of all neurons in the ganglion cell layer are amacrine cells, whereas in the peripheral temporal, upper and lower retina, the number of displaced amacrine cells exceeds the number of ganglion cells by a factor of five. In peripheral nasal retina, however, there are always more ganglion cells than displaced amacrine cells.

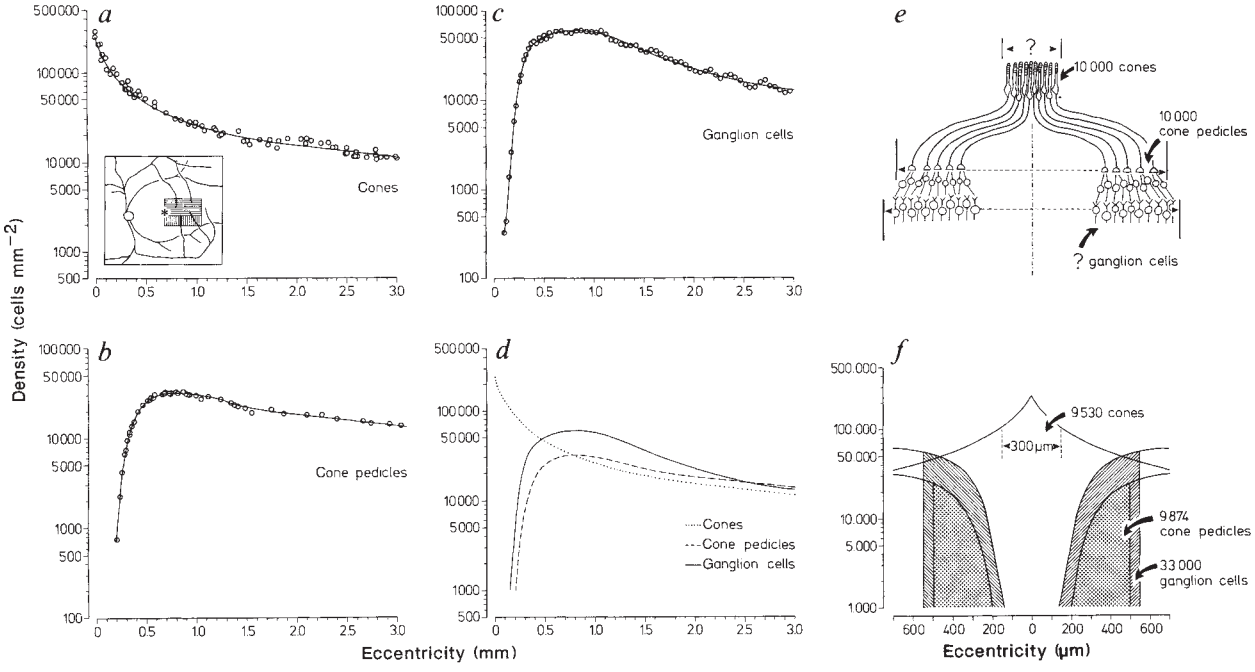


FIG. 2 Topography of the macaque monkey fovea. The inset in *a* shows the retinal area examined. The optic-nerve head (open circle), the fovea (asterisk) and some blood vessels characterize the central retina. *a*, *b*, and *c*, Cone inner segment, cone pedicle and ganglion cell densities, measured in temporal retina from the foveal centre up to 3 mm eccentricity. For the first 800 μm the curves show counts from the two sectioned retinae. In *d*, the densities of cones, cone pedicles and ganglion cells are compared. The data presented in *a*–*f* are not corrected for shrinkage. The ganglion cell to cone ratio is not affected by shrinkage, because all counts were from the same material. *e*, Schematic drawing of the monkey fovea. From the cone inner segments in the centre of the fovea, cone pedicles are displaced radially by up to 350 μm owing to the Henle fibres. In an annular zone the cone pedicles make contact with bipolar cells, which in turn synapse onto ganglion cells. *f*, Density curves of cones, cone pedicles and ganglion cells in the monkey fovea. The curves are derived from *d*, assuming radial symmetry for the foveal centre. Nasal and temporal densities did not differ within the foveal centre. The cone densities were integrated within the central 300 μm (9,530 cones), the cone pedicles were integrated in a circle around the centre of 500 μm radius (dotted area, 9,874 cone pedicles) and ganglion cells were integrated within a circle of 550 μm radius (hatched area, 33,000 ganglion cells).

METHODS. A rectangle (3.5 × 2.5 mm) was cut from the eye cup, embedded in Epon and serially sectioned at 1 μm thickness. Vertical sections (perpendicular to the retinal layers) were cut through the upper two thirds of the block, and horizontal sections (parallel to the retinal layers) through the lower one third. From the other eye of the same monkey a comparable block (2 × 1 mm) centred at the fovea was dissected out and serially sectioned vertically. Sections were stained with toluidine blue, dehydrated and mounted in Permount. Within the first 0.8 mm from the fovea, cone inner segments and cone pedicles were counted from the vertical sections of either eye; from 0.6–3 mm they were counted in horizontal sections. Ganglion cells were counted from vertical sections. The rest of the retinae were processed as whole mounts, and cone and ganglion cell densities obtained for the large eccentricities. The direct counts from the whole mounts and horizontal sections were also used to calculate a correction factor for the densities measured from the vertical sections[27]. This result was cross-checked by the disector method[28,29]. All three measurements gave results that differed by less than 10%.

Ganglion cell axons innervate relay cells in the lateral geniculate, where they remain separated in different layers corresponding to the eye of origin[9]. In the visual cortex (V1), information from the nasal retina of one eye and the temporal retina of the other are combined. The foveal representation comprises the sum of ganglion cells from both eyes (1,336,500 cells mm^{-2}). But because the nasal retina is wider than the temporal retina, at the large eccentricities only the contralateral eye is represented in the visual cortex (ganglion cell density, 800 cells mm^{-2}). If it is assumed that there is equal cortical space per ganglion cell, then it can be predicted that between the centre and the periphery the cortical representation changes by a factor of 1,670. This is in good agreement with the results of physiological studies in which 1,000–4,000-fold changes of the areal cortical-magnification factor have been reported[1-8]. On this basis it is not necessary to postulate an additional magnification in the lateral geniculate and/or the visual cortex[7,9-11]. So far, a more precise comparison between the ganglion cell density and the cortical magnification factor cannot be made on the basis of published data. The physiological measurements provide a linear magnification factor (mm of cortex per degree—often measured along only one visual axis), from which the areal magnification factor cannot be obtained if the cortical magnification is anisotropic along different axes[22,23].

In the cat visual system there is a close correspondence between retinal ganglion cell density and the decrease of the areal cortical magnification factor with eccentricity[24]. In the central area of the cat retina, the sampling density of ON-β cells is 2,000–3,000 mm^{-2}, and the cortical magnification factor is 1–4 mm^2 $degree^{-2}$ (refs 24, 25). In the monkey fovea the sampling density of ON-Pβ cells is 267,300 mm^{-2}, and the cortical magnification factor is ~ 100–300 mm^2 $degree^{-2}$ (ref. 8). Both the ganglion cell density and the areal magnification factor are 100 times those of the cat. It seems that the rules of mapping the retina onto the visual cortex are similar.

Distorted representations in the cortex are also found for other sensory modalities; the 'homunculus' of the somatotopic map on the human cortex[31] is probably the best-known example. The simple rule—that the central representation follows the peripheral receptor or neural density—may be applicable not only to the visual system but to other sensory modalities. □

Received 24 July; accepted 7 September 1989.

1. Talbot, S. A. & Marshall, W. H. *Am. J. Ophthalmol.* **24,** 1255–1264 (1941).
2. Daniel, P. M. & Whitteridge, D. *J. Physiol., Lond.* **159,** 203–211 (1961).
3. Cowey, A. *J. Neurophysiol.* **27,** 366–393 (1964).
4. Rolls, E. T. & Cowey, A. *Expl Brain Res.* **10,** 298–310 (1970).
5. Hubel, D. H. & Wiesel, T. N. *J. comp. Neurol.* **158,** 295–305 (1974).
6. Dow, B. M., Snyder, A. Z., Vautin, R. G. & Bauer, R. *Expl Brain Res.* **44,** 213–228 (1981).
7. Van Essen, D. C., Newsome, W. T. & Maunsell, J. H. R. *Vision Res.* **24,** 429–448 (1984).
8. Tootell, R. B. H., Switkes, E., Silverman, M. S. & Hamilton, S. L. *J. Neurosci.* **8,** 1531–1568 (1988).
9. Malpeli, J. G. & Baker, F. H. *J. comp. Neurol.* **161,** 569–594 (1975).
10. Myerson, J., Marris, P. B., Miezin, F. M. & Allman, J. M. *Science* **198,** 855–857 (1977).
11. Perry, V. H. & Cowey, A. *Vision Res.* **25,** 1795–1810 (1985).
12. Drasdo, N. *Nature* **266,** 554–556 (1977).
13. Rovamo, J. & Virsu, V. *Expl Brain Res.* **37,** 495–510 (1979).
14. Schein, S. J. & de Monasterio, F. M. *J. Neurosci.* **7,** 996–1009 (1987).
15. Wong, R. O. L. & Hughes, A. *J. comp. Neurol.* **255,** 159–177 (1987).
16. Wässle, H., Chun, M. H. & Müller, F. *J. comp. Neurol.* **265,** 391–408 (1987).
17. Missotten, L. *Invest. Opthalmol. Vis. Sci.* **13,** 1045–1049 (1974).
18. Perry, V. H. & Cowey, A. *Neuroscience* **25,** 225–236 (1988).
19. Schein, S. J. *J. comp. Neurol.* **269,** 479–505 (1988).
20. Boycott, B. B., Hopkins, J. M. & Sperling, H. G. *Proc. R. Soc.* B **229,** 345–379 (1987).
21. Röhrenbeck, J., Wässle, H. & Boycott, B. B. *Eur. J. Neurosci.* **1,** 407–420 (1989).
22. Sakitt, B. *Vision Res.* **22,** 417–421 (1982).
23. Mallot, H. A. *Biol. Cybern.* **52,** 45–51 (1985).
24. Tusa, R. J., Palmer, L. A. & Rosenquist, A. C. *J. comp. Neurol.* **177,** 213–235 (1978).
25. Sanderson, K. J. *Expl Brain Res.* **13,** 159–177 (1971).
26. Wässle, H. & Chun, M. H. *J. comp. Neurol.* **297,** 43–54 (1989).
27. Abercrombie, M. *Anat. Rec.* **94,** 239–247 (1946).
28. Sterio, D. C. *J. Microsc.* **134,** 127–136 (1984).
29. Gundersen, H. J. G. *J. Microsc.* **143,** 3–45 (1986).
30. Allen, K. A., Curcio, C. A. & Kalina, R. E. *Invest. Ophthalmol. Vis. Sci.* **30,** 347 (1989).
31. Penfield, W. & Rasmussen, T. *The Cerebral Cortex of Man: A Clinical Study of Localization of Function* (Macmillan, New York, 1950).
32. Koontz, M. A., Hendrickson, A. E. & Ryan, M. K. *Vis. Neurosci.* **2,** 19–25 (1989).

ACKNOWLEDGEMENTS. We thank Heide Ahmed, Felicitas Boij, Barbara Tuschen and Walter Hofer for technical assistance, Irmgard Odenthal for secretarial services, Ulf Eysel for making the photocoagulator lesions, and Paul Martin for helpful discussions.

Development of glomerular pattern visualized in the olfactory bulbs of living mice

Anthony-Samuel LaMantia & Dale Purves

Department of Anatomy and Neurobiology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, Missouri 63110, USA

MANY regions of the mammalian brain are characterized by iterated ensembles of nerve cells which can be distinguished anatomically and physiologically. A particularly striking example is the pattern of glomeruli in the olfactory bulbs[1-4]; other instances are columns and 'blobs' in the visual cortex[5], barrels and columns in the somatosensory cortex[6,7], and striasomes and cell islands in the neostriatum[8,9]. Understanding the generation of these neuronal ensembles has a bearing on a variety of important neurobiological problems, including the nature of critical periods, the age-dependent response of the nervous system to injury and the manner in which neural information is stored. Analysis of these issues has usually been restricted to studies of the brains of different individuals at various ages. Many questions about the formation of such units, however, can only be answered by observing the same brain repeatedly in a living animal. This strategy would enable a direct assessment of how these units are assembled, whether the initial ensembles persist and whether units are lost or gained as an animal matures. We have succeeded in studying the pattern of glomeruli in the mouse olfactory bulb on two separate occasions during postnatal development. Comparison of the patterns observed at intervals of up to three weeks show that this part of the brain is gradually constructed by the addition of new glomeruli to a persisting population.

Glomeruli are discrete spheres of neuropile found in the outer layer of the olfactory bulbs of mammals and many other vertebrates[1,2] (Fig. 1*a*); electrophysiological and anatomical studies have shown that these units perform the initial processing of olfactory information[3,4]. To monitor their development, vital staining techniques used to explore the dynamic properties of individual nerve cells and synapses in the peripheral nervous system[10] have been adapted for use in the central nervous system. Visualizing the living brain over time, however, presents special problems. First, many salient features of neural organization are located well beneath the brain surface. Second, the dural, arachnoid and pial membranes must remain intact to preserve the integrity of the brain during the interval between observations. These membranes limit the access of vital dyes to the underlying neural elements and present a substantial impediment to microscopical imaging. Such difficulties are minimized in the mouse olfactory bulb, where the glomeruli are near the dorsal surface of the brain (Fig. 1*a*, *b*) and where the overlying membranes are relatively thin.

To observe directly how the mature pattern of glomeruli develops, 4–6-day-old mice (CF-1 strain, either sex, 2.5–3.5 g) were anaesthetized by hypothermia[11] and placed on the stage of a scanning confocal microscope[12,13] adapted for *in vivo* imaging[10,14]. To visualize glomeruli, the right bulb was exposed through a window made in the overlying bone (Fig. 1*b*), and the diffusion barrier presented by the meninges breached transiently by treatment with a hypertonic saline solution[15]. The glomeruli were then stained by applying a dilute solution of the styryl dye RH-414[16-18] to the intact dura. Within several minutes, a pattern of discrete fluorescent spheres appeared within the superficial portion of the bulb (Fig. 1*c*). Images of the fluorescently stained bulb were then digitized and stored in computer disk files. Routine histological staining (Fig. 1*d*) confirmed that each of the fluorescent spheres in these images corresponds to a glomerulus (see Fig. 2). Once the *in vivo* images had been obtained, the pups were allowed to recover and returned to the

litter after ~30 min. At intervals of 1 h to 3 weeks, a permanent histological record of the final arrangement of glomeruli was made for comparison with the initial *in vivo* images.

Images obtained in the same animal after an interval of ~1 h showed identical patterns of glomeruli (Fig. 2*a*). In eight mice examined, a total of 446 glomeruli were observed in the living state. All of these glomeruli had the same size, shape and arrangement in the final histological preparation. We saw no evidence of cryptic or nascent glomeruli that were observable with one technique but not the other. These results demonstrate that the initial visualization of glomeruli in the living animal is both complete and reliable.

At longer intervals the addition of new glomeruli was apparent. Comparison of images obtained initially and after one week showed that all of the glomeruli observed in the first image (262 glomeruli in 7 mice) could be identified in the second. However, new glomeruli were also seen, representing an average increment of 14.5% over this period. There was no evidence of glomerular elimination. After 2–3 weeks (Fig. 2*b*), the glomeruli initially observed (329 glomeruli in 9 animals) were still present. At this interval, the degree of glomerular addition had increased further, the average increment being 28.3%. The magnitude of addition of new glomeruli for each of the animals studied at these several intervals is shown in Fig. 3. This rate of addition is consistent with glomerular counts in fixed material in both mice[19] and rats[20]. Our *in vivo* observations show that new glomeruli are gradually added to a persisting population of these functional units.

Examination by light and electron microscope showed no consistent differences between the imaged and the contralateral bulbs, either acutely or after intervals of 2–3 weeks. There are, however, some inherent uncertainties in our approach. First, we cannot monitor the pattern of glomeruli continuously. Therefore, we must infer that a glomerulus in a similar position, having a similar shape, and with similar neighbour relationships at a second observation is, in fact, the same glomerulus initially observed in that location. Second, as new glomeruli are added and individual glomeruli grow larger, the relationships among the original population become progressively distorted (see Fig. 2). It is, therefore, increasingly difficult to compare the initial

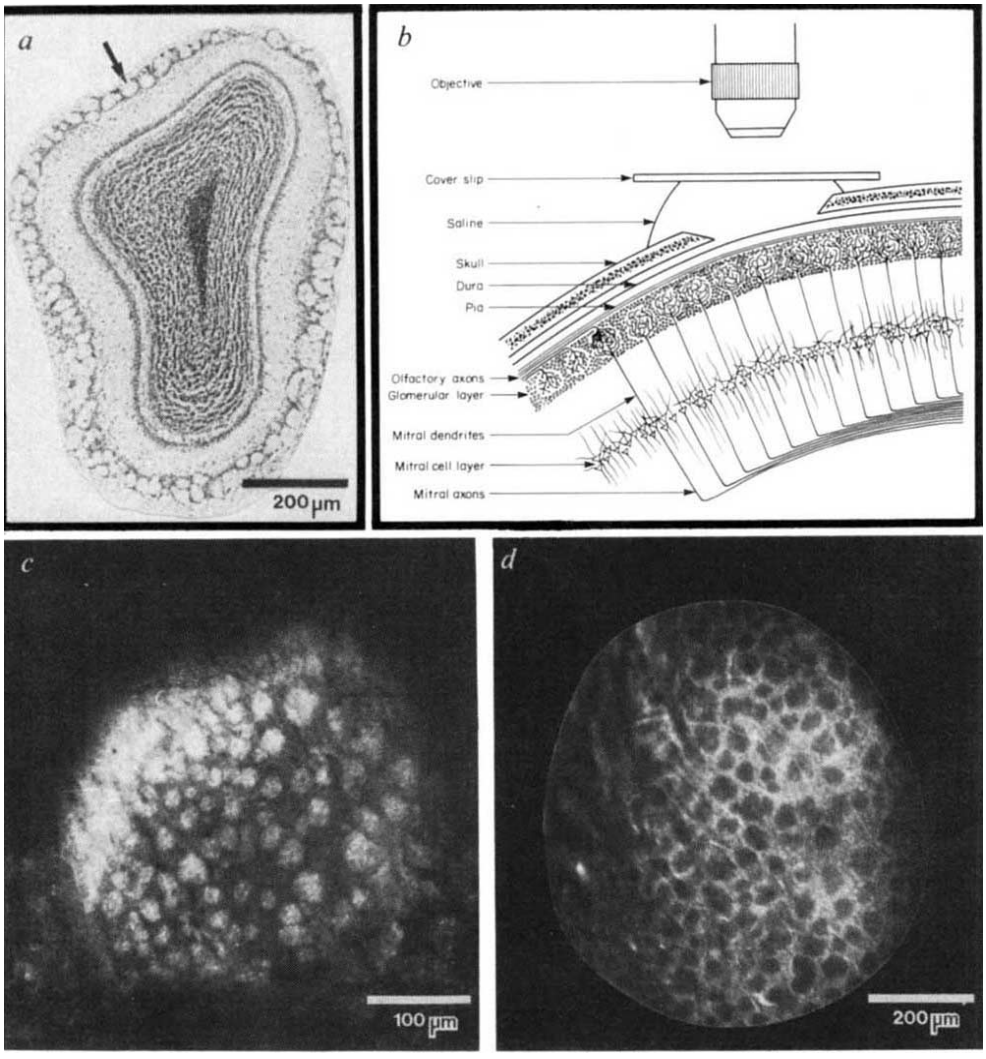


FIG. 1 Methods of visualizing olfactory glomeruli. *a*, Frontal section (50 μm) through the right olfactory bulb of a four-week-old mouse after aldehyde fixation and staining with cresyl violet (dorsal surface is up, lateral to the left). The outer layer of the mammalian bulb (arrow) comprises spherical units of neuropile called glomeruli, the borders of which are defined by a continuous rind of closely packed periglomerular cells. In the adult mouse olfactory bulb, these glomeruli number ~2,000 and, on average, measure 75 μm in diameter (S. L. Pomeroy, A-S.L, and D.P., unpublished data). The glomeruli on the dorsal surface of the bulb are generally arranged in a single layer that lies just beneath a thin mantle of afferent olfactory axons. *b*, Diagram of the surgical approach used to visualize olfactory glomeruli in the intact animal (orientation same as *a*). *c*, Confocal video image of the dorsal surface of a five-day-old mouse olfactory bulb obtained after fluorescence staining *in situ*; the bright spots are the olfactory glomeruli. In this view, anterior is up and lateral to the left. The image was obtained with an MRC LaserSharp-500 scanning confocal microscope (BIORAD), which was used to reduce out-of-focus fluorescence arising from the overlying afferent fibre layer and the meninges. The light source is an argon ion laser, fitted with 488 nm excitation and 515 nm emission filters; for each image, 25 frames were averaged using a 4× Planapo lens (numerical aperture = 0.1) at a scan rate of 1 frame per second. The depth of focus at this magnification is sufficient to image the glomerular pattern in a single plane. The laser setting was kept on low power and the intensity further reduced by a 10% transmittance neutral density filter. To stain the glomeruli, the meningeal diffusion barrier was first breached by a 3 min pretreatment with 1.6 M saline; this step was immediately followed by a 3 min exposure to a 1.5 mg ml^{-1} solution of RH-414 (Molecular Probes, Eugene, Oregon). Based on the staining characteristics of styryl dyes in the periphery[30–32], and previous descriptions of olfactory receptor axons[2,4,33,34], we surmise that many of the vitally stained elements within glomeruli are the presynaptic terminals of olfactory receptor axons. Control experiments showed that the fluorescent staining of glomeruli faded within a few hours and was not apparent 1–2 days later. *d*, Photomicrograph of the dorsal surface of the olfactory bulb stained with Sudan Black. Mice were anaesthetized with nembutal (50 mg kg^{-1}), perfused transcardially with aldehyde fixatives, and the bulbs stained '*en bloc*'. The entire bulb was immersed in a 0.7% solution of Sudan Black in propylene glycol for 10 min, differentiated by washing in propylene glycol and water and then mounted whole in gelatin. In this way we could obtain a permanent record of the glomerular pattern at any interval after the acquisition of an initial *in vivo* image.

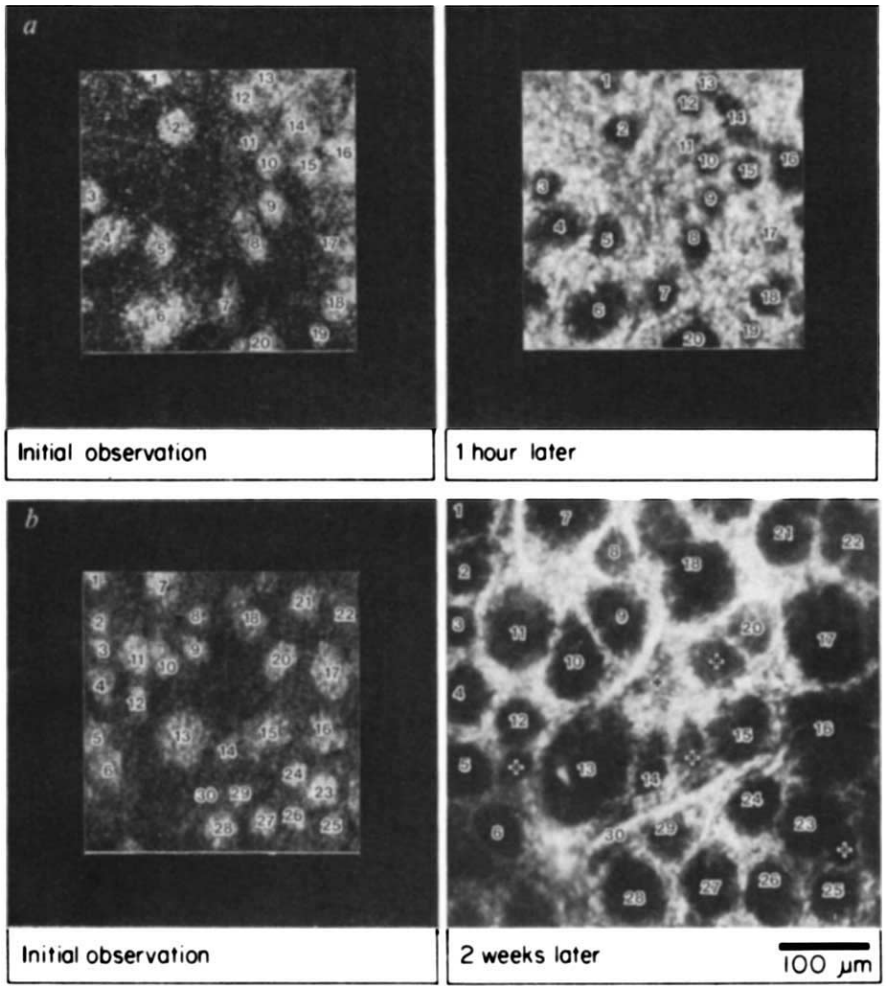


FIG. 2 Representative examples of the pattern of olfactory glomeruli observed in 4–6-day-old mice initially and after an interval of *a*, 1 h, or *b*, 2 weeks. *a* and *b* represent different animals; in each case, the left panel shows the initial image obtained *in vivo*, and the right panel the final pattern observed after staining the fixed bulb with Sudan Black. Numbers indicate corresponding glomeruli; +, indicates a glomerulus added during the interval between observations. New glomeruli were typically small compared with members of the original population; note also the substantial growth of glomeruli over the 2-week interval.

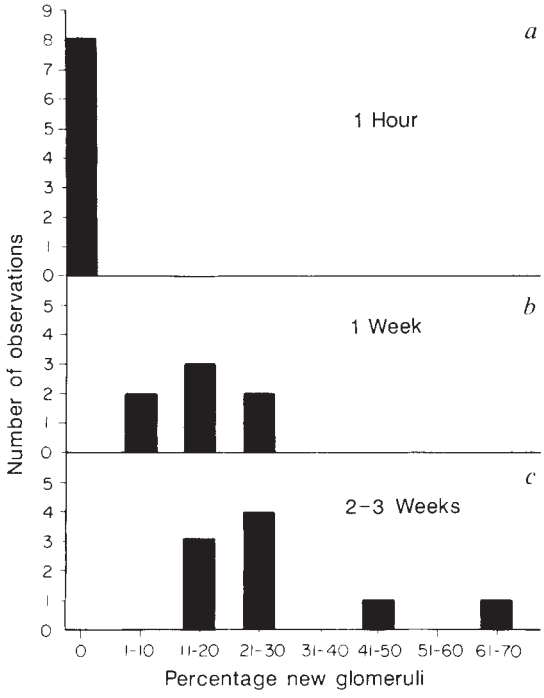


FIG. 3 Magnitude of addition of new glomeruli after intervals of *a*, 1 h; *b*, 1 week; or *c*, 2–3 weeks. Each observation represents a separate animal in which the number of new glomeruli observed in the final image is expressed as a percentage of the number of original glomeruli (all of which were retained; see Fig. 2).

and final maps with confidence at intervals of more than 3 weeks.

By monitoring the same region of the developing brain in a particular animal, we have shown that the full complement of neural circuits is not present at birth; instead, the architecture of the mouse olfactory bulb is established gradually during postnatal life by the addition of new functional units, without apparent deletions from the original population. These observations do not discount the important contribution of cell death, synaptic loss and axon elimination to the development of the brain[21-24]; they show, however, that such regressive phenomena, if they occur in this region, proceed during the ongoing construction of neural circuits. The progressive elaboration of functional units might be exaggerated in the olfactory system, which has several extraordinary features[25,26]. We note, however, that the postnatal growth of the rest of the mouse brain is approximately the same as that of the olfactory bulbs (S. Pomeroy, A.-S.L. and D.P., unpublished data). Evidence that development of at least this part of the mouse central nervous system involves ongoing construction must now be reconciled with theories that argue that brain development proceeds by the selection of useful circuits from a larger initial repertoire[27-29]. □

Received 28 July; accepted 14 September 1989.

1. Golgi, C. *Riv. Sper. Freniat. Med. leg Alien Ment.* **1,** 405–425 (1874).
2. Ramón y Cajal, S. in *Gaceta Sanitaria de Barcelona* 133–139, 174–181, 206–212 (1890).
3. Shepherd, G. M. *Physiol. Rev.* **53,** 864–917 (1972).
4. Scott, J. W. & Harrison, T. A. in *Neurobiology of Taste and Smell* (eds Finger, T. E. & Silver, W. L.) 151–178 (Wiley, New York, 1987).
5. Hubel, D. H. *Eye, Brain and Vision.* (Scientific American Library, New York, 1988).
6. Woolsey, T. A. & Van der Loos, H. *Brain Res.* **17,** 205–242 (1970).
7. Mountcastle, V. B. in *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function* (eds Edelman, G. M. & Mountcastle, V. B.) 7–50 (MIT, Cambridge, Massachusetts, 1978).

8. Goldman-Rakic, P. S. *J. comp. Neurol.* **205,** 398–413 (1982).
9. Graybiel, A. M. & Ragsdale, C. W. *Proc. natn. Acad. Sci. U.S.A.* **75,** 5723–5726 (1978).
10. Purves, D. & Voyvodic, J. *Trends Neurosci.* **10,** 398–404 (1987).
11. Green, C. J. *Animal Anaesthesia* (Laboratory Animals Ltd, London, 1979).
12. White, J. G., Amos, W. B. & Fordham, M. *J. Cell Biol.* **105,** 41–48 (1987).
13. Fine, A., Amos, W. B., Durbin, R. M. & McNaughton, P. A. *Trends Neurosci.* **11,** 346–351 (1988).
14. Purves, D., Hadley, R. D. & Voyvodic, J. *J. Neurosci.* **6,** 1051–1060 (1986).
15. Fenstermacher, J. D. & Rapoport, S. I. in *Handbook of Physiology. The Cardiovascular System IV* Vol. IV (eds Renkin, E. M. & Michel, C. C.) 969–1000 (American Physiological Society, Bethesda, 1984).
16. Grinvald, A., Anglister, L., Freeman, J. A., Hildesheim, R. & Manker, A. *Nature* **308,** 848–850 (1984).
17. Orbach, H. S., Cohen, L. B. & Grinvald, A. *J. Neurosci.* **5,** 1886 (1985).
18. Kauer, J. S. *Nature* **331,** 166–168 (1988).
19. Pomeroy, S. L., LaMantia, A.-S. & Purves, D. *Soc. Neurosci. Abstr.* **15,** 806 (1989).
20. Meisami, E. & Safari, L. *Brain Res.* **221,** 81–107 (1981).
21. Oppenheim, R. W. in *Studies in Developmental Neurobiology. Essays in Honor of Viktor Hamburger* (ed. Cowan, W. M.) 74–133 (Oxford University Press, 1981).
22. Hamburger, V. & Oppenheim, R. W. *Neurosci. Comment.* **1,** 39–55 (1982).
23. Purves, D. & Lichtman, J. W. *Science* **210,** 153–157 (1980).
24. Cowan, W. M., Fawcett, J. W., O'Leary, D. D. M. & Stanfield, B. B. *Science* **225,** 1258–1265 (1984).
25. Hinds, J. W. *J. comp. Neurol.* **134,** 287–304 (1967).
26. Graziadei, P. P. C. & Monti-Graziadei, G. A. *J. Neurocytol.* **8,** 197–213 (1979).
27. Changeux, J.-P. & Danchin, A. *Nature* **264,** 705–712 (1976).
28. Young, J. Z. *J. Roy. Soc. Med.* **72,** 801–814 (1979).
29. Edelman, G. M. *Neural Darwinism: the Theory of Neuronal Group Selection* (Basic Books, New York, 1987).
30. Magrassi, L., Purves, D. & Lichtman, J. W. *J. Neurosci.* **7,** 1207–1214 (1987).
31. Purves, D., Voyvodic, J. T., Magrassi, L. & Yawo, H. *Science* **258,** 1122–1126 (1987).
32. Balice-Gordon, R. & Lichtman, J. W. *Soc. Neurosci. Abstr.* **15,** 165 (1989).
33. Pinching, A. J. & Powell, T. P. S. *J. Cell Science* **9,** 347–377 (1971).
34. Land, L. S., Eager, R. P. & Shepherd, G. M. *Brain Res.* **25,** 250–254 (1970).

ACKNOWLEDGEMENTS. We are grateful to Ladan Hedayati, Chris Ho, Pat Newton, Scott Pomeroy and David Simon for advice and assistance, and to Amarim Grinvald for gifts of RH-414 and other fluorescent dyes.

Positive selection of CD4$^+$T cells mediated by MHC class II-bearing stromal cell in the thymic cortex

Jerome Bill* & Ed Palmer*†‡

* Division of Basic Sciences, Department of Pediatrics, National Jewish Center for Immunology and Respiratory Medicine, 1400 Jackson Street, Denver, Colorado 80262, USA
† Department of Microbiology and Immunology, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA

T LYMPHOCYTES differentiate in the thymus, where functionally immature, CD4$^+$CD8$^+$ (double positive) thymocytes develop into functionally mature CD4$^+$ helper cells and CD8$^+$ cytotoxic (single positive) T cells. The thymus is the site where self-reactive T cells are negatively selected (clonally deleted)[1–10] and where T cells with the capacity to recognize foreign antigens in association with self-proteins encoded by the major histocompatibility complex (MHC) are positively selected[11]. The net result of these developmental pathways is a T-cell repertoire that is both self-tolerant and self-restricted. One unresolved issue is the identity of the thymic stromal cells that mediate the negative and positive selection of the T-cell repertoire. Previous work has pointed to a bone-marrow-derived macrophage or dendritic cell as the inducer of tolerance[12,13], whereas a radiation-resistant[14,15], deoxyguanosine-resistant[16] thymic cell seems to mediate the positive selection of self-MHC restricted T cells. Thymic stromal cells in the cortex interact with the T-cell antigen receptor on thymocytes[17]. Using several strains of transgenic mice that express the class II MHC molecule I–E in specific regions of the thymus[18], we show directly that the positive selection of T cells is mediated by an I–E-bearing cell in the thymic cortex.

Previous experiments by MacDonald *et al.*[19] and from our laboratory (data not shown) have demonstrated that I-E expression mediates positive selection of CD4$^+$ Vβ6$^+$ thymocytes. To directly examine which thymic stromal cells are responsible for the elevated levels of Vβ6$^+$ T cells in I-E-bearing mice, we have determined the frequency of T cells expressing the Vβ6 domain in several strains of transgenic mice[18,20]. The Eα16 strain (I-Ewt) is transgenic for the $E\alpha^k$ gene and expresses the *I-E* gene product on stromal cells in the thymic cortex, thymic medulla and on B cells and macrophages. The ΔX21-16 strain (ΔX) carries an $E\alpha^k$ transgene from which the 14-base pair (bp) 'X' box in the 5′ upstream region has been deleted. In this strain there is no detectable I-E expression in the thymic cortex. There is, however, normal expression of the I-E protein on stromal cells in the thymic medulla and on B cells and macrophages. The 14-bp 'Y' box in the 5′ upstream region has been deleted from the $E\alpha^k$ transgene present in the ΔY301-54(ΔY) strain. Except for a small number of dendritic cells, I-E expression is absent from the thymic medulla of this strain. In ΔY mice, I-E is expressed on stromal cells in the thymic cortex and on 20–50% of peripheral B cells, but not on macrophages. Finally, the Sma58 strain (Sma) is transgenic for an $E\alpha^k$ gene from which a 1-kilobase (kb) fragment has been deleted from the 5′ upstream 'B-cell control' region. The I-E antigen is expressed normally in the thymic medulla and cortex but is absent from 98% of the B cells in this strain. All of these transgenes have been bred onto a B10.S(7R).βL genetic background[21]. This strain carries C57BL/10 background genes, the $H\text{-}2^{s(7R)}$ MHC (K^s, $I\text{-}A^s$, $E\beta^s$, $E\alpha^-$, D^d), and the T-cell antigen receptor (TCR) β-chain genes from C57L mice. This β-chain allele contains the Vβ1, 2, 3, 4, 6, 7, 10, 14, 15, 16 and 17a gene segments.

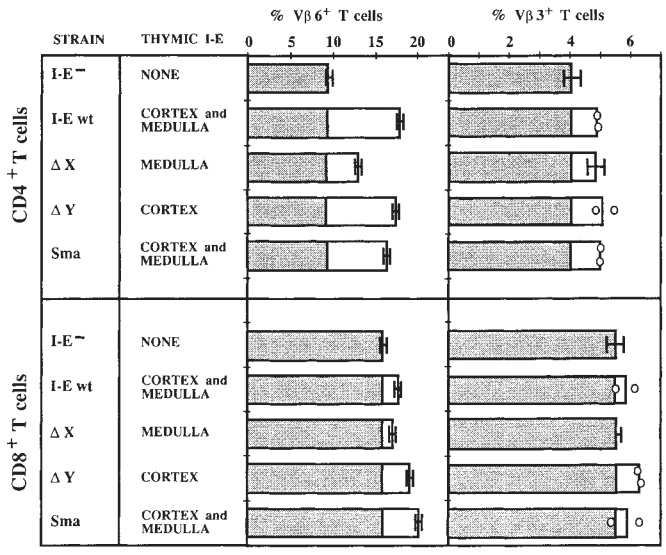


FIG. 1 Flow cytometric analysis of Vβ6-bearing peripheral lymph node T cells in transgenic mice that express I-E on various cell types. Double staining of CD4$^+$ or CD8$^+$ and Vβ6 or Vβ3 T cells is shown. Stippled portions of bars indicate the levels of Vβ6$^+$ or Vβ3$^+$ T-cells in transgenic negative (I-E$^-$) mice. I-E$^-$ mice are non-transgenic littermates.
METHODS. Lymph-node cells that were non-adherent to nylon wool were doubly stained with biotinylated RR4-7 (anti-Vβ6; ref 4) or KJ25 (anti-Vβ3; ref. 5) followed by phycoerythrin-conjugated streptavidin (Tago) and with FITC-GK1.5 (anti-CD4; ref. 28) or FITC-53.6 (anti-CD8; ref. 29) and analysed on an EPICS C or EPICS 751 flow cytometer (Coulter Electronics). Bars indicate average values from two or more individual mice; s.e.m. is given when 3 mice or more were studied, otherwise individual values are indicated by dots. The Eα 16 (I-Ewt), ΔX21-16, ΔY301-54 and Sma58 transgenes were first back-crossed onto SJL as described[18], and for this study were also back-crossed onto the C57BL/10 background by breeding to B10.S(7R)βL mice. The B.10.S(7R)βL strain carries a deletion at the TCR β-locus, which reduces the number of *Vβ* gene segments by half (refs 21, 30, 31). See text for a description of the I-E phenotypes of these strains. The increased CD4$^+$ Vβ6$^+$ level in ΔY mice compared with that in I-E$^-$ and ΔX mice is statistically significant ($P<0.00005$). The increased CD8$^+$ Vβ6$^+$ level in ΔY mice compared with that in I-E$^-$ and ΔX mice are also significant ($P<0.00005$ and $P<0.0006$ respectively). Statistical significance was determined by analysis of variance with contrast using SAS (Cary).

‡ To whom correspondence should be addressed.

The data in Fig. 1 show the frequency of Vβ6-bearing T cells in these strains. Vβ6$^+$ cells account for only 9.7% of the CD4$^+$ T lymphocytes in B10.S(7R).βL (I-E$^-$) mice (non-transgenic littermates); on the other hand, 17.9% of the CD4$^+$ T cells are Vβ6$^+$ in B10.S(7R).βL mice expressing the wild-type, $E\alpha^k$ gene (*I-E*wt). The ΔY strain expresses wild-type levels of Vβ6$^+$, CD4$^+$ peripheral T cells, even though I-E expression is virtually absent from the thymic medulla, from 50–80% of the B cells, and from nearly all of the macrophages in these mice. Immunohistological staining reveals that I-E expression is normal in the thymic cortex of ΔY mice[18]. The ΔX mice do not carry wild-type levels of Vβ6$^+$, CD4$^+$ peripheral T cells and do not express the *I-E* gene product in the thymic cortex. The Sma strain which expresses I-E normally in the thymus carries the wild-type level of Vβ6$^+$, CD4$^+$ peripheral T cells, even though I-E is not expressed on macrophages and is expressed only on 2% of peripheral B cells. Because Sma mice express wild-type levels of Vβ6$^+$ T cells, we suggest that I-E-bearing cells in the periphery have little or no role in the positive selection of the T-cell repertoire. Taken together, these results indicate that the I-E-directed, positive selection of Vβ6-bearing T cells is mediated by an I-E-expressing cell in the thymic cortex. The data in Fig. 1 show that the frequency of Vβ6$^+$, CD4$^+$ T cells in ΔX mice is increased over that seen in I-E$^-$ mice. It is conceivable that a few stromal cells in the thymic medulla have the capacity to positively select developing thymocytes, although there may not be enough of these cells to fully select the repertoire. Alternatively, large numbers of CD4$^+$ CD8$^+$ (double positive) thymocytes, residing in the thymic cortex, may not be able physically to interact with I-E-bearing cells in the medulla of ΔX mice. We would not expect a class II MHC molecule to affect the positive selection of CD8$^+$ T cells[10,22–25], but there is a significant if slight increase in the frequency of CD8$^+$, Vβ6$^+$ T cells in transgenic mice expressing I-E in the thymic cortex. These cells probably represent Vβ6$^+$ lymphocytes that have been positively selected by I-E but which inappropriately express CD8. A comparable 'inappropriate' expression of CD4 was found by Scott *et al.*[26] on 15% of idiotype$^+$ T cells in female SCID mice transgenic for a class I-restricted anti-H-Y *TCR*.

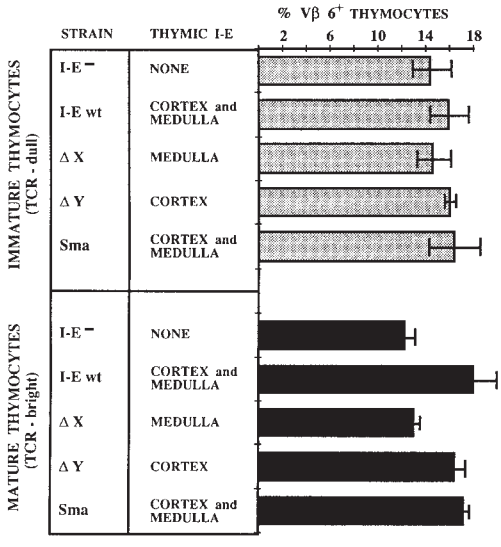


FIG. 2 Flow cytometric analysis of Vβ6 expression in I-E transgenic mice and I-E-negative littermates. Immature thymocytes (TCR-dull), stippled bars; mature thymocytes (TCR-bright), solid bars.

METHODS. Mice are described in the legend to Fig. 1. Before staining, thymocytes were cultured in complete tumour medium to increase the density of TCR on the immature population[27]. The percentage of total thymocytes displaying the immature (low-density TCR) or mature (high-density TCR) phenotypes was determined by parallel staining with H57-597, which detects all murine αβ TCRs[31,32]. Staining used biotinylated H57-597 or RR4-7 followed by phycoerythrin-streptavidin and cells were analysed as described in Fig. 1. The increased Vβ6$^+$ level in the mature thymocytes of ΔY mice over I-E$^-$ mice is statistically significant ($P<0.02$). Statistical significance was determined by analysis of variance with contrast using SAS (Cary).

It has been previously reported that Vβ17a-bearing cells are clonally deleted from I-Ewt, ΔX, ΔY and Sma mice to a similar extent. By introducing I-E expression into these animals, the level of Vβ17a$^+$ T cells decreases from 9% to 3% of the total number of T cells[18]. As the number of T cells is decreased by at least 6% in I-E-bearing animals, the fraction of all other (Vβ-bearing) T cells increases proportionately. To monitor the magnitude of this effect, we have examined the frequency of Vβ3-bearing T cells in these I-E transgenic strains (Fig. 1). Unlike the dramatic increase in Vβ6$^+$, CD4$^+$ T cells, the increase in Vβ3$^+$ T cells in I-E-bearing animals is barely detectable and is the same, regardless of the site of I-E expression in the thymus. This stands in contrast to the positive selection of Vβ6-bearing T cells, which requires I-E expression in the thymic cortex.

The data in Fig. 2 show the fraction of Vβ6-bearing thymocytes in these I-E transgenic strains. The fraction of Vβ6$^+$ thymocytes has been determined in both the TCR-dull (functionally immature, CD4$^+$ and CD8$^+$) and TCR-bright (functionally mature, CD4$^+$ or CD8$^+$) thymocyte populations[27]. The frequency of Vβ6-bearing cells among immature thymocytes is similar among all the strains we examined, regardless of whether or where the I-E gene product is expressed. Conversely, only those strains that express I-E in the thymic cortex (I-Ewt, ΔY, Sma) have an increased frequency of mature—CD4$^+$ or CD8$^+$ (single positive)—thymocytes. In agreement with the observations of others[10,23], the effects of positive selection can be seen in the mature but not the immature thymocyte population. These data are also consistent with the idea that the positive selection of thymocytes is related to the differentiation of TCR-dull (CD4$^+$ CD8$^+$-double positive) into TCR-bright (CD4$^+$ or CD8$^+$-single positive) cells.

These experiments help to identify the thymic stromal-cell type responsible for the positive selection of self-MHC-restricted T lymphocytes. Previous work has indicated that a radiation-resistant[14,15], deoxyguanosine-resistant[16] thymic cell mediates the positive selection of self-MHC-restricted T cells. Our results demonstrate that an I-E-bearing cell in the thymic cortex is important for the positive selection of CD4$^+$ T cells bearing the Vβ6 domain. The most abundant cell type in the thymic cortex is the epithelial cell, which apparently interacts with immature thymocytes to select those with specificity for self-MHC and foreign antigens. A clarification of the genetic and physiological properties of stromal cells in the thymic cortex that distinguish them from other cell types may shed light on the cellular interactions involved in the selection of a self-MHC-restricted T-cell repertoire. □

Received 30 May; accepted 1 September 1989.

1. Kappler, J. W., Roehm, N. & Marrack, P. *Cell* **49,** 273–280 (1987).
2. Kappler, J. W., Staerz, U., White, J. & Marrack, P. *Nature* **262,** 797–799 (1988).
3. MacDonald, H. R. *et al. Nature* **332,** 40–45 (1988).
4. Kanagawa, O., Palmer, E. & Bill, *J. Cellular Immunology* **119,** 412–426 (1989).
5. Pullen, A. M., Marrack, P. & Kappler, J. W. *Nature* **335,** 796–801 (1988).
6. Abe, R., Vacchio, M. S., Fox, B. & Hodes, R. J. *Nature* **335,** 827–830 (1988).
7. Tomonari, K. & Lovering, E. *Immunogenetics* **28,** 445–451 (1988).
8. Bill, J., Kanagawa, O., Woodland, D. L. & Palmer, E. *J. exp. Med.* **169,** 1405–1419 (1989).
9. Kisielow, P., Bluthmann, H., Staerz, U. D., Steinmetz, M. & von Boehmer, H. *Nature* **333,** 742–746 (1988).
10. Sha, W. C. *et al. Nature* **336,** 73–76 (1988).
11. Zinkernagel, R. & Doherty, P. *Nature* **248,** 701–702 (1974).
12. Ready, A. R., Jenkinson, E. J., Kingston, R. & Owen, J. J. T. *Nature* **310,** 231–233 (1984).
13. von Boehmer, H. & Schubiger, K. *Eur. J. Immun.* **14,** 1048–1052 (1983).
14. Zinkernagel, R. M. *et al. J. exp. Med.* **147,** 883–897 (1978).
15. Bevan M. & Fink, P. *Immun. Rev.* **42,** 3–19 (1978).
16. Lo, D. & Sprent, J. *Nature* **319,** 672–675 (1986).
17. Farr, A. G., Anderson, S. K., Marrack, P. & Kappler, J. *Cell* **43,** 543–550 (1985).
18. van Ewijk, W. *et al. Cell* **53,** 357–370 (1988).
19. MacDonald, H. R., Lees, R. K., Schneider, R., Zinkernagel, R. M. & Hengartner, H. *Nature* **336,** 471–473 (1988).
20. Dorn, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **84,** 6249–6253 (1987).
21. Kappler, J. W., Kushnir, E. & Marrack, P. *J. exp. Med.* **169,** 1533–1541 (1989).
22. Sprent, J., Schaefer, M., Lo, D. & Korngold, R. *Immun. Rev.* **91,** 195–218 (1986).
23. Kisielow, P., Teh, H. S., Bluthmann, H. & von Boehmer, H. *Nature* **335,** 730–733 (1988).

24. Marusic-Galesic, S., Stephany, D. A., Longo, D. L. & Kruisbeek, A. M. *Nature* **333**, 180–183 (1988).
25. Zuniga-Pflucker, J. C., Longo, D. L. & Kruisbeek, A. M. *Nature* **338**, 76–78 (1989).
26. Scott, B., Blüthmann, H., Teh, H. S. & von Boehmer, H. *Nature* **338**, 591–593 (1989).
27. Roehm, N. *et al. Cell* **38**, 577–584 (1984).
28. Dialynas, D. *et al. J. Immun.* **131**, 2445–2451 (1983).
29. Ledbetter, J. & Herzenberg, L. *Immun. Rev.* **47**, 63–90 (1979).
30. Chou, H. S., Nelson, C. A., Godambe, S. A., Chaplin, D. D. & Loh, D. Y. *Science* **238**, 545–548 (1987).
31. Wilson, R. K., Lai, E., Concannon, P., Barth, R. K. & Hood, L. E. *Immun. Rev.* **101**, 149–172 (1988).
32. Kubo, R. T., Born, W., Kappler, J. W., Marrack, P. & Pigeon, M. *J. Immun.* **142**, 2736–2742 (1989).

ACKNOWLEDGEMENTS. We are indebted to Drs Diane Mathis and Christophe Benoist for providing the I-Ewt, ΔX, ΔY and Sma mice. We also thank Drs P. Marrack, J. Kappler and Ms Ella Kushnir for breeding the B10.S(7R).βL mice, Drs Brian Kotzin and John La Breque for help with statistical analysis and Dr Mary Pat Happ for critical review of the manuscript. The KJ25mAb was obtained from Drs A. Pullen, J. Kappler and P. Marrack, H57-597 mAb from Dr Ralph Kubo. Ed Palmer is a recipient of a Faculty Research Award from the American Cancer Society. This work was supported in part by the NIH.

Intrathymic signalling in immature CD4⁺CD8⁺ thymocytes results in tyrosine phosphorylation of the T-cell receptor zeta chain

Toshinori Nakayama*, Alfred Singer*, Eric D. Hsi† & Lawrence E. Samelson†

* Experimental Immunology Branch, National Cancer Institute, and † Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

THYMIC selection of the developing T-cell repertoire occurs in immature CD4⁺CD8⁺ double-positive thymocytes[1–3] and is thought to be mediated by signals transduced by T-cell antigen receptor (TCR) molecules and possibly by CD4 and CD8 accessory molecules as well[1,4–6]. It is not known, however, which signal-transduction mechanisms function in immature CD4⁺CD8⁺ thymocytes on engagement of TCR, CD4 or CD8 molecules. In mature T cells, CD4 and CD8 molecules are each associated with the *src*-like protein tyrosine kinase p56 *lck*[7,8] and signals transduced by TCR and CD4 activate tyrosine kinases that phosphorylate TCR-ζ chains and other intracellular substrates[9,10]. Consequently, we examined whether tyrosine kinases could be similarly activated in immature CD4⁺CD8⁺ thymocytes. Unexpectedly, we found that TCR-ζ chains from CD4⁺CD8⁺ thymocytes were already phosphorylated *in vivo*, and that dephosphorylation of this TCR subunit occurred on removal of CD4⁺CD8⁺ cells from their intrathymic environment. Rephosphorylation of TCR-ζ in cultured CD4⁺CD8⁺ thymocytes occurred rapidly *in vitro*, either in response to cross-linking of TCR, CD4 or CD8 by specific monoclonal antibodies, or on cell–cell contact. These observations indicate that tyrosine kinases are activated *in vivo* in immature CD4⁺CD8⁺ thymocytes undergoing thymic differentiation and selection. They also indicate that TCR, CD4 and CD8 molecules can function in CD4⁺CD8⁺ thymocytes as signalling molecules to activate tyrosine kinases and that phosphorylated TCR-ζ serves as a marker of these signalling events.

We began our studies by examining tyrosine phosphorylation levels of TCR-ζ in young-adult thymocytes. Tyrosine phosphorylation of TCR-ζ can be readily detected by immunoprecipitating intact antigen receptors from cell lysates with anti-TCR monoclonal antibodies, and then immunoblotting with affinity purified phosphotyrosine-specific antibodies[10]. We were surprised to find that TCR-ζ chains in freshly isolated thymocytes were already tyrosine phosphorylated (Fig. 1*a*, *b*), including TCR-ζ from immature CD4⁺CD8⁺ thymocytes (Fig. 1*c*). This latter result was obtained by immunoprecipitation with anti-Vβ17a monoclonal antibody which, in C57BR mice, binds only a small subset of immature CD4⁺CD8⁺ cells[11], and so a longer exposure time is required for detection than for C57L samples, where the anti-Vβ17a antibody binds to other thymocytes in addition to CD4⁺CD8⁺ cells[11]. The phosphoprotein observed in these immunoblots had a relative molecular mass (M_r) of 21,000 (21K) and was clearly TCR-ζ because it was directly precipitated by anti-ζ antibodies, and because it migrated under non-reducing conditions with an apparent molecular mass of 32K, a characteristic of TCR-ζ[12]. Also, the immunoprecipitation by anti-TCR monoclonal antibodies of the 21K phosphoprotein was only accomplished by anti-TCR antibodies appropriate to the strain from which the thymocytes were obtained (anti-CD3 and anti-Vβ8 monoclonal antibodies in C57BL/6 thymocytes, but not anti-Vβ17a; anti-CD3 and anti-Vβ17a antibodies in C57BR and C57L thymocytes, but not anti-Vβ8)[13].

The apparent constitutive phosphorylation of TCR-ζ in freshly isolated thymocytes indicated either that receptor-mediated tyrosine kinase activation had already occurred in these thymocytes *in vivo*, or that ζ-chain tyrosine phosphorylation had occurred for other unknown reasons. Attempts to stimulate additional thymocyte ζ-chain phosphorylation by TCR cross-linking *in vitro* failed (data not shown), indicating that phosphorylation was either already maximal or could not be induced by TCR in these cells. We then examined the effect of removing

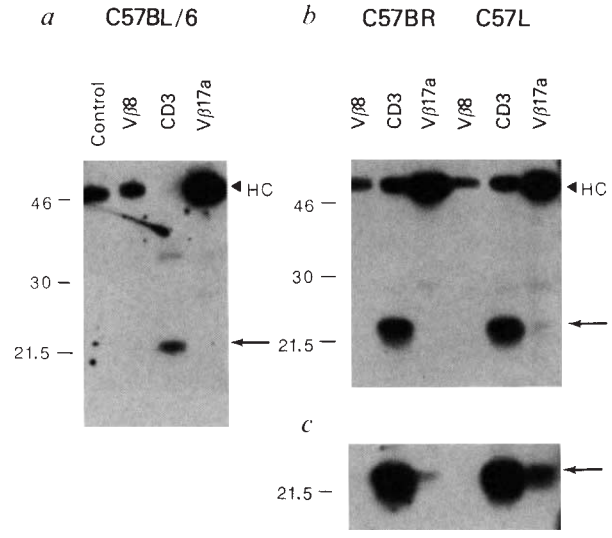


FIG. 1 Phosphorylation of the TCR-ζ chain in freshly isolated murine thymocytes. TCR from normal mice were isolated by immunoprecipitation and electrophoresis in SDS-PAGE, and tyrosine-phosphorylated TCR-ζ detected by immunoblotting with anti-phosphotyrosine antibodies. *a*, Thymocytes from C57BL/6 (H-2^b, Vβ8⁺, Vβ17a⁻) mice; *b* and *c*, thymocytes from C57BR/CDJ (H-2^k, Vβ8⁻, Vβ17a⁺) and C57L (H-2^b, Vβ8⁻, Vβ17a⁺) mice. The position of phosphorylated TCR-ζ chain is indicated by arrows; immunoglobulin heavy chains (HC) are also noted. Relative molecular mass markers (left) are (from the top of the gel): ovabumin (46K), carbonic anhydrase (30K), and trypsin inhibitor (21.5K). Low levels of tryosine-phosphorylated CD3-ε are seen in lanes with anti-CD3 immunoprecipitates. The 28K band immunoprecipitated with KJ23a appears to be nonspecific as it is seen in Vβ17a⁻ cells (panel *a*).

METHODS. Thymocytes from normal adult mice were isolated, washed and solubilized to maintain intact, phosphorylated TCR–CD3 complex by a protocol using Triton X-100, protease and phosphatase inhibitors[20]. The concentration of EDTA (5 mM) in the lysis buffer is sufficient to inhibit artefactual tyrosine phosphorylation that might occur during the isolation procedure[10]. Immunoprecipitation of the TCR was performed with monoclonal antibodies absorbed to protein-A agarose. The antibodies used were: a negative control recognizing a clonotypic determinant expressed on a single T-cell clone (A2B4.2), anti-Vβ8 (F23.1) (ref. 21), anti-CD3-ε (145-2C11) (ref. 22), and anti-Vβ17a (KJ23a) (ref. 11). After immunoprecipitation, electrophoresis on a 13% acrylamide gel and electrotransfer, phosphorylated TCR-ζ chain was detected by immunoblotting with affinity-purified anti-phosphotyrosine antibodies and ^{125}I-labelled protein-A. The nitrocellulose membrane was exposed to Kodak XAR-2 film for 2 days (panels *a* and *b*) or for 2 weeks (panel *c*) at −70 °C.

thymocytes from their intrathymic environment by culturing them briefly *in vitro* at 37 °C, and observed a marked dephosphorylation of TCR-ζ (Fig. 2*a*). This dephosphorylation of TCR-ζ did not occur at 4 °C and did not reflect a general decrease in tyrosine phosphorylation of intracellular substrates induced during the short-term culture, as the amount of tyrosine phosphorylation of other substrates did not decrease (Fig. 2*a*). Consistent, therefore, with receptor disengagement and kinase inactivation, short-term culture of thymocytes resulted in preferential dephosphorylation of TCR-ζ.

We reasoned that if phosphorylation of TCR-ζ chains in freshly isolated thymocytes reflected intrathymic receptor-mediated signalling events, re-engagement of surface receptors should transduce signals that would induce the rephosphorylation of TCR-ζ. To test this prediction, we cultured thymocytes at 37 °C for various times and stimulated them for the final 30 min with anti-CD3 monoclonal antibody in the presence of Fc-receptor (FcR)-positive LK35.2 (LK) tumour cells, to induce multivalent antibody cross-linking. As seen in Table 1, TCR-ζ dephosphorylation was significant after only one hour of culture. Similarly, detection of CD3-induced rephosphorylation required only one to two hours in culture. The rapidity of these *in vitro* events indicates that TCR signalling of tyrosine kinase activation can occur in thymocytes without further differentiation and suggests that removal from the thymus results in receptor disengagement and receptor resensitization. Indeed, it is conceivable that phosphorylation of the TCR-ζ chain itself negatively regulates TCR signalling.

To assess the ability of other cell-surface molecules to induce TCR-ζ rephosphorylation, we cultured thymocytes at 37 °C overnight, and then stimulated them for 30 min with various monoclonal antibodies in the presence of FcR^+ LK cells (Fig. 2*b*, *c*). Anti-$\alpha\beta$TCR, anti-CD3, anti-CD4, and anti-CD8 antibodies each induced 4–7-fold increases in ζ-chain phosphorylation in whole thymocyte populations as compared with thymocytes stimulated by medium or LK cells alone, whereas anti-H-2K and anti-CD45 antibodies were either non-stimulatory or only minimally stimulatory (Fig. 2*b*, *c*). Evidence that receptor-mediated rephosphorylation of TCR-ζ specifically occurred in immature $CD4^+CD8^+$ thymocytes was obtained using anti-TCR reagents that selectively bind $CD4^+CD8^+$ cells ultimately destined for clonal elimination[11,14–17]. Stimulation of C57BR thymocytes with anti-Vβ17a[11], and stimulation of CBA/J thymocytes with anti-Vβ6[15] and anti-Vβ3 antibodies[16,17] resulted in small but significant (1.5–2-fold) increases in ζ-chain phosphorylation (Fig. 2*b*, *c*), consistent with the antibodies binding only a small subset of $CD4^+CD8^+$ thymocytes in these strains.

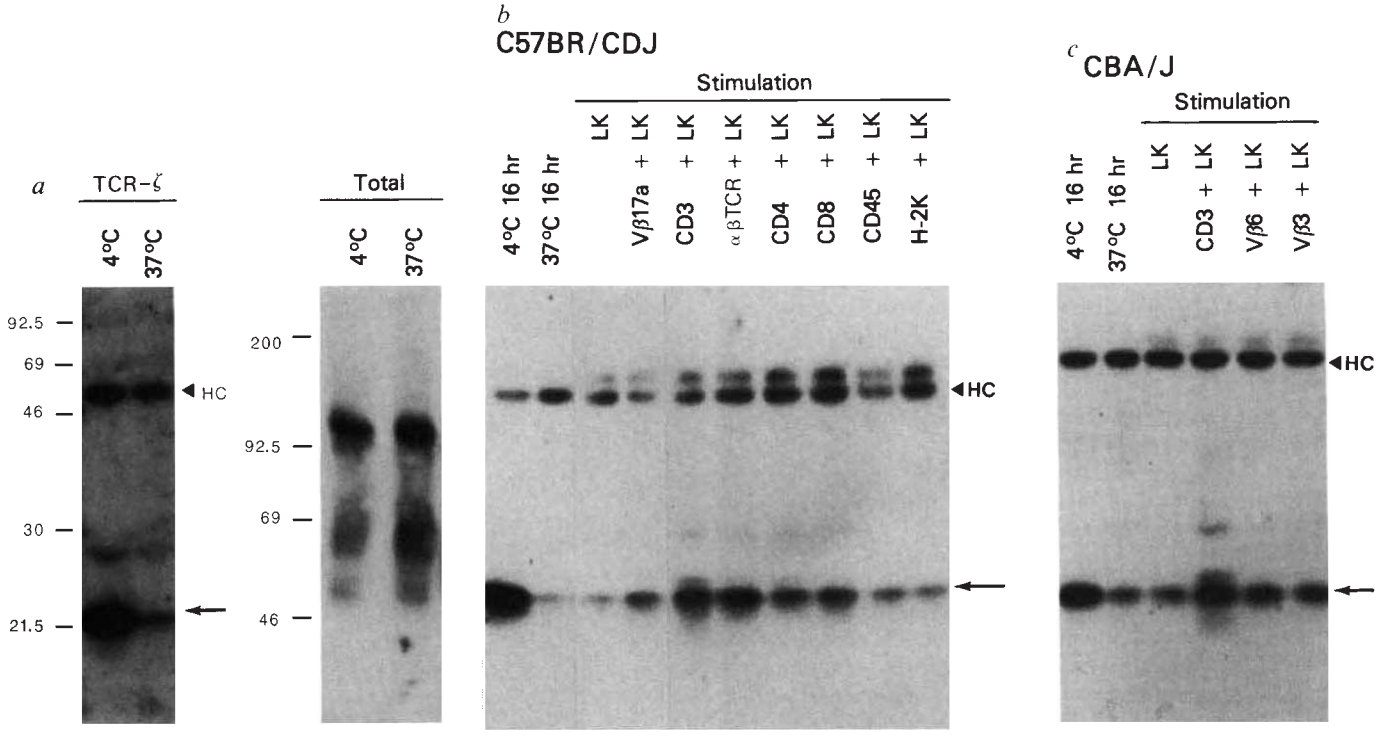


FIG. 2 Dephosphorylation and rephosphorylation of TCR-ζ chain in cultured murine thymocytes. *a*, C57BL/10 thymocytes were cultured for 16 h at either 4 °C or 37 °C. Phosphorylated TCR-ζ (left) or total tyrosine-phosphorylated molecules (right) were detected by anti-phosphotryosine immunoblotting. The level of TCR-ζ phosphorylation detected in thymocytes cultured for 16 h at 4 °C was equal to that of freshly isolated thymocytes (data not shown). Relative molecular mass markers (K) are shown on the left of each gel. *b*, Thymocytes from C57BR/CDJ (H-2^k) were cultured for 16 h at 37 °C and then stimulated for 30 min with monoclonal antibodies and a multivalent cross-linker in the form of Fc-receptor-(FcR) positive LK35.2 (LK) tumour cells[23]. The antibodies used were: anti-TCR Vβ17a (KJ23a), anti-CD3 (145-2C11), anti-$\alpha\beta$TCR (H57-597) (ref. 24), anti-CD4 (GK1.5) (ref. 25), anti-CD8 (53-6.72) (ref. 26), anti-CD45 (M1/9) (ref. 27), and anti-H-$2K^k$ (11-4.1). The TCR-CD3 complex from the thymocytes was immunoprecipitated with anti-CD3-ε, (145-2C11). The phosphorylated TCR-ζ (arrow) was detected by immunoblotting using phosphotyrosine-specific antiserum. *c*, Thymocytes from CBA/J (H-2^k, $Mls^{a,c}$) were cultured for 16 h at 37 °C and then stimulated for 30 min with monoclonal antibodies and FcR^+ LK cells. The antibodies used were: anti-CD3 (145-2C11), anti-TCR Vβ6 monoclonal antibody (RR4-7) (ref. 28), and anti-TCR Vβ3 (KJ25a) (ref. 16).

METHODS. Normal adult thymocytes (1×10^7) were cultured in 24-well culture plate (Coster 3524) in RPMI 1640 supplemented with 10% fetal calf serum at either 4 °C or at 37 °C for 16 h. In *a* (left), phosphorylated TCR-ζ was detected by immunoblotting using anti-phosphotyrosine antibodies after immunoprecipitation with anti-CD3 monoclonal antibody (145-2C11) as described in Fig. 1. Tyrosine-phosphorylated substrates in the cultured thymocytes were detected by electrophoresis of the post-nuclear lysate from 5×10^6 cells. For *b* and *c*, cultured cells (37 °C for 16 h) were stimulated with culture supernanant of monoclonal antibodies (25% v/v final concentration) in the presence of FcR^+ LK cells (1×10^7 ml^{-1}) at a 2:1 ratio of thymocytes to tumour cells, at 37 °C for 30 min. The TCR-CD3 complexes (3×10^7 thymocytes equivalent) were immunoprecipitated with anti-CD3 (145-2C11) and subjected to SDS-PAGE, immunoblotting and autoradiography as described in Fig. 1. Exposure times for panels *a*, *b* and *c* were 3 days, 2 days and 1 day, respectively. Relative band intensities were determined by densitometry.

To assess signalling in immature $CD4^+CD8^+$ thymocytes directly, we selected $CD4^+CD8^+$ thymocytes from whole thymus populations, obtaining cell populations that were >96% $CD4^+CD8^+$. The purified thymocytes were cultured overnight to permit TCR-ζ chain dephosphorylation, and then stimulated for 30 min with anti-CD3, anti-CD4, and anti-CD8 monoclonal antibodies in the presence of FcR^+ LK cells as a multivalent cross-linker (Fig. 3*a*). All three antibodies induced significant (2.5–3-fold) rephosphorylation of TCR-ζ (Fig. 3*a*). Control add-back experiments confirmed that the TCR-ζ rephosphorylation observed in Fig. 3*a* occurred in immature $CD4^+CD8^+$ thymocytes, and not simply in contaminating mature single-positive cells (data not shown). Thus, CD3, CD4, and CD8 cell-surface molecules are all coupled to a functional tyrosine kinase pathway in immature $CD4^+CD8^+$ thymocytes that, upon activation, phosphorylates TCR-ζ.

We next determined whether tyrosine kinase activation could occur in immature $CD4^+CD8^+$ thymocytes by direct cell–cell interactions in the absence of antibodies. LK tumour cells, which we had used in previous experiments simply as a source of FcR^+ cells during short-term stimulation with monoclonal antibodies, bear multiple MHC class I and class II molecules that are the natural ligands for CD4, CD8, and TCR. Consequently, we added LK cells for longer periods to cultures of $CD4^+CD8^+$ thymocytes in the absence of monoclonal antibodies and assessed their effect on TCR-ζ phosphorylation (Fig. 3*b*). The TCR-ζ chains from $CD4^+CD8^+$ thymocytes were far more extensively phosphorylated when LK cells were present throughout the 16-hour culture period than if they were not (Fig. 3*b*, compare lanes 2 and 3). LK cells, therefore, either inhibited the dephosphorylation of TCR-ζ in $CD4^+CD8^+$ thymocytes, which otherwise occurs during 16 hours *in vitro*, or induced TCR-ζ rephosphorylation, or both. To determine the ability of LK cells to induce ζ-chain rephosphorylation in immature thymocytes, we dephosphorylated ζ-chains, culturing for 8 hours before the addition of LK cells for an additional 8 hours (Fig. 3*b*). It can be seen that 8-hour stimulation of pre-cultured $CD4^+CD8^+$ thymocytes with LK cells induced a 3–3.5-fold increase in ζ-chain rephosphorylation (compare lanes 2, 4 and 5). Thus, cell–cell interactions, in the absence of monoclonal antibodies, can activate the tyrosine-phosphorylation signalling pathway in immature $CD4^+CD8^+$ thymocytes.

Previous studies of signalling in immature $CD4^+CD8^+$ thymocytes measured calcium transients induced by TCR engagement and found them to be present, but blunted[18,19]. The results presented here indicate that these blunted responses might be due to prior *in vivo* receptor engagement and receptor desensitization. This study also identifies activation of tyrosine kinases as a functional signalling pathway in immature $CD4^+CD8^+$ thymocytes triggered by $\alpha\beta$TCR, CD3, CD4, or CD8 molecules.

TABLE 1 Kinetics of ζ-chain phosphorylation

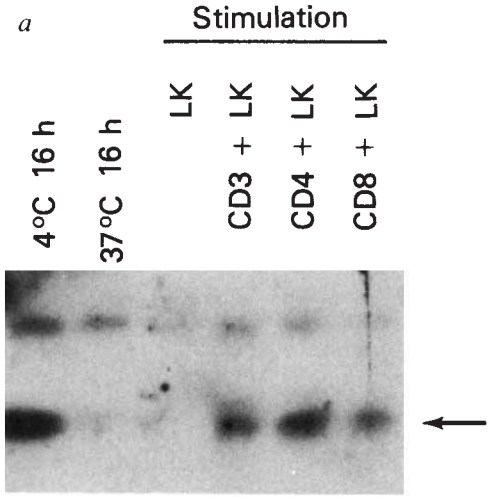
Time in culture (h)	ζ-Chain tyrosine phosphate* Unstimulated thymocytes†	Anti-CD3-stimulated thymocytes‡	Ratio of stimulated/unstimulated thymocytes
	Optical density§		
0	2.0	ND	
1	1.2	1.4	1.2
2	1.1	2.4	2.2
3.5	0.5	1.8	3.6

* C57BL/6 thymocytes (5×10^7) were solubilized in detergent either immediately or at the indicated times after culture at 37 °C. Intact phosphorylated TCR–CD3 complex was immunoprecipitated with 145–2C11 and subjected to SDS-PAGE, immunoblotting and autoradiography, as described in the legend to Fig. 1.

† Thymocytes (5×10^7) cultured in medium alone for the indicated time period.

‡ Thymocytes (5×10^7) were cultured in medium alone and then, for the final 30 min of culture, co-cultured with 145–2C11 culture supernatant (4% v/v final concentration) in the presence of FcR^+ LK cells at a 2:1 ratio of thymocytes to tumour cells. Total time in culture, including the 30 min anti-CD3 stimulation, was as indicated.

§ Arbitrary densitometric units.



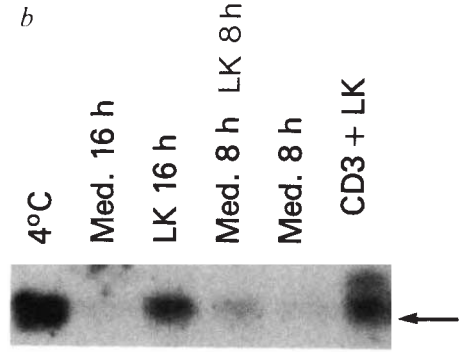


FIG. 3 Receptor-mediated signalling of tyrosine kinase activated in cultured $CD4^+CD8^+$ thymocytes. *a*, $CD4^+CD8^+$ enriched thymocytes from C57BL/6 mice were cultured at 4 °C or 37 °C for 16 h and then stimulated for 30 min with monoclonal antibodies and FcR^+ LK cells. The antibodies used were: anti-CD3-ε (145–2C11), anti-CD4 (GK1.5), and anti-CD8 (53–6.72). The tyrosine phosphorylation of TCR-ζ was detected by anti-phosphotyrosine antibodies (arrow). *b*, $CD4^+CD8^+$ enriched thymocytes from C57BL/6 mice were cultured at 37 °C for 16 h either in medium alone or together with the B-cell line, LK. Where indicated, thymocytes were cultured in media alone for 8 h and then co-cultured with LK for an additional 8 h. The phosphorylation stimulated by anti-CD3 +LK for the last 30 min of culture was also assessed as a positive control. Phosphorylated ζ-chain is indicated by an arrow. Results are representative of 3 experiments.

METHODS. $CD4^+CD8^+$ thymocytes from C57BL/6 (H-2^b, Lyt2.2) mice were enriched by applying thymocytes to anti-Lyt-2.2 (83-12-5) (ref. 29)-coated dishes (Falcon 1029) for 90 min at 4 °C. Adherent cells were collected and subjected to another cycle of anti-Lyt-2.2 panning. The anti-Lyt-2.2 monoclonal antibody used in the panning does not come off the plate and so does not interfere with subsequent staining of the cells with anti-CD8 (data not shown). The purity of the adherent cells was tested by two-colour staining using an anti-CD4 antibody (H129.19) (ref. 30) conjugated with fluorescein isothiocyanate (FITC) and a biotinylated anti-CD8 (53.6-72), followed by incubation with Texas-red-avidin.[6] The purity of $CD4^+CD8^+$ cells was >96%, with <1% contamination by $CD4^+CD8^-$ cells and <3% contamination by $CD4^-CD8^+$ cells. In *b*, 1×10^7 purified thymocytes were co-cultured with 5×10^6 LK cells in 24-well culture plate. The stimulations, immunoprecipitations and immunoblotting were performed as described in Figs 1 and 2. Exposure time was 4 days for panel *a* and 9 days for panel *b*. Relative band intensities were determined by densitometry.

Finally, we show that tyrosine kinase activation occurs *in vivo* in immature $CD4^+CD8^+$ thymocytes undergoing differentiation and selection, and that phosphorylated TCR-ζ serves as a marker of these signalling events. □

Received 17 July; accepted 23 August 1989.

1. Kisielow, P., Bluthmann, H., Staerz, U. D., Steinmetz, M. & von Boehmer H. *Nature* **333,** 742-746 (1988).
2. Fowlkes, B. J., Schwartz, R. H. & Pardoll, D. M. *Nature* **334,** 620-623 (1988).
3. Sha, W. C. *et al. Nature* **336,** 73-76 (1988).
4. Eichmann, K., Jonsson, J.-I., Falk, I. & Emmrich, F. *Eur. J. Immun.* **17,** 643-650 (1987).
5. Mizuochi, T., Tentori, L., Sharrow, S. O., Kruisbeek, A. M. & Singer, A. *J. exp. med.* **168,** 437-442 (1988).
6. McCarthy, S. A., Kruisbeek, A. M., Uppenkamp, I. K., Sharrow, S. O. & Singer, A. *Nature* **336,** 76-79 (1988).
7. Veillette, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. *Cell* **55,** 301-308 (1988).
8. Rudd, C. E., Trevillyan, J. M., Dasgupta, J. D., Wong, L. L. & Schlossman, S. F. *Proc. natn. Acad. Sci. U.S.A.* **85,** 5190-5194 (1988).
9. Veillette, A., Bookman, M. A., Horak, E. M., Samelson, L. E. & Bolen, J. B. *Nature* **338,** 257-259 (1989).
10. Hsi, E. D. *et al. J. biol. Chem.* **264,** 10836-10842 (1989).
11. Kappler, J. W., Roehm, N. & Marrack, P. *Cell* **49,** 273-280 (1987).
12. Baniyash, M., Garcia-Morales, P., Luong, E., Samelson, L. E. & Klausner, R. D. *J. biol. Chem.* **263,** 18225-18230 (1988).
13. Kappler, J. W. *et al. Cell* **49,** 263-271 (1987).
14. Kappler, J. W., Staerz, U., White, J. & Marrack, P. C. *Nature* **332,** 35-40 (1988).
15. MacDonald, H. R. *et al. Nature* **332,** 40-45 (1988).
16. Pullen, A. M., Marrack, P. & Kappler, J. W. *Nature* **335,** 796-801 (1988).
17. Abe, R., Vacchio, M. S., Fox, B. & Hodes, R. *Nature* **335,** 827-830 (1988).
18. Havran, W. L. *et al. Nature* **330,** 170-173 (1987).
19. Finkel, T. H., McDuffie, M., Kappler, J. W., Marrack, P. & Cambier, J. C. *Nature* **330,** 179-181 (1987).
20. Samelson, L. E., Patel, M. D., Weissman, A. M., Harford, J. B. & Klausner, R. D. *Cell* **46,** 1083-1090 (1986).
21. Staerz, U. D., Rammensee, H.-G., Benedetto, J. D. & Bevan, M. J. *J. Immun.* **134,** 3994-4000 (1985).
22. Leo, O., Foo, M., Sachs, D. H., Samelson, L. E. & Bluestone, J. A. *Proc. natn. Acad. Sci. U.S.A.* **84,** 1374-1378 (1987).
23. Kappler, J., White, J., Weggman, D., Mustain, E. & Marrack, P. *Proc. natn. Acad. Sci. U.S.A.* **79,** 3604-3607 (1982).
24. Kubo, R. T., Born, W., Kappler, J. W., Marrack, P. & Pigeon, M. *J. Immun.* **142,** 2736-2742 (1989).
25. Dialynas, D. P. *et al. Immunol. Rev.* **74,** 29-56 (1983).
26. Ledbetter, J. A. & Herzenberg, L. A. *Immunol. Rev.* **47,** 63-90 (1979).
27. Springer, T., Galfre, G., Secher, D. S. & Milstein, C. *Eur. J. Immun.* **8,** 539-551 (1978).
28. Kanagawa, O., Palmer, E. & Bill, J. *Cellular Immun.* **119,** 421-426 (1989).
29. Leo, O., Foo, M., Segal, D. M., Shevach, E., & Bluestone, J. A. *J. Immun.* **139,** 1214-1222 (1987).
30. Pierres, A. *et al. J. Immun.* **132,** 2775-2782 (1984).

ACKNOWLEDGEMENTS. We are grateful to Dr Richard Klausner and Dr Richard Hodes for discussion and critical reading of the manuscript. E.D. Hsi was a Howard Hughes Medical Institute Scholar.

Identification of a novel retinoic acid receptor in regenerative tissues of the newt

Clifton W. Ragsdale Jr*, Martin Petkovich†, Phillip B. Gates*, Pierre Chambon† & Jeremy P. Brockes*

* The Ludwig Institute for Cancer Research, Courtauld Building, 91 Riding House Street, London W1P 8BT, UK
† Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médicine, 11 rue Humann, 67085 Strasbourg Cedex, France

IN urodele amphibians, the progenitor cells that regenerate amputated limbs (known as the blastema) normally replace only the missing structures[1]. After systemic delivery of retinoic acid (RA), more proximal structures are also formed, indicating that RA can control position specification in the proximal–distal axis of the regenerating limb[2–4]. According to dose and experimental context, retinoids can also re-specify the anteroposterior axis of the limb, induce deletions of skeletal elements, or block re-growth completely[2,5–10]. To study the molecular basis of these morphogenetic effects, we screened complementary DNA libraries of newt regenerative tissues (limbs and tails) for hormone nuclear receptors activated by RA[11–17]. Two functional retinoic acid receptors (RARs) were identified, one of which is the newt homologue of the human α-receptor (RARα). The second receptor, called RARδ, is novel. Sequence analysis suggests that the composite newt RAR previously reported[18] is chimaeric, consisting of 5′RAR-β-like and 3′ RARδ clones. We conclude that multiple RARs are expressed during limb regeneration in amphibians and suggest that receptor heterogeneity may underlie the different effects of retinoids on limb morphogenesis.

To identify RARs that could participate in amphibian limb regeneration, we screened an amplified newt-forelimb blastema λgt11 cDNA library with a human RARα probe[11] under low-stringency hybridization conditions. Fifty plaques out of a population of 1.2×10^6 gave positive signals and five of these were isolated and sequenced. Two of the clones, λE2 and λE3, identified the putative newt homologue of the mammalian RARα. Figure 1*a* presents the nucleotide sequence and amino-acid translation of the longest open reading frame (ORF) in λE3, which predicts a protein of 458 amino acids of relative molecular mass (M_r) of 50,638. Table 1 records the amino-acid sequence identity of the λE3 cDNA translation with that reported for the 462-amino-acid human RARα. A high degree of sequence conservation between the newt and human receptors is seen across all regions of the molecule, including regions A and F. Of particular note is that 24 of the 26 amino acids at the C-terminal end of the receptor are identical between the species, indicating that this part of the F region makes an important contribution to RAR function.

Three clones (λE1, λE4 and λE6) identified the C-terminal half of a second RAR (Fig. 1*b*). Two of these clones, λE1 and λE4, although of different size (1.5 and 3.2 kilobases (Kb), respectively), share 5′ ends, which suggests that *Eco*RI sites native to the cDNA sequence may have been cut during cloning. To obtain a 5′ extension of the sequence, we used a 164-bp *Eco*RI-*Sph*I fragment from the λE1 clone to screen a random-primed λZAP cDNA library of normal newt tail that was prepared with *Eco*RI adaptors. A single phage, λNRO, was isolated and sequenced. This cDNA overlapped the λE1 clone by 407 bp (all identical) and provided a 5′ 660-bp extension. RNase protection experiments using a 299-bp *Hin*dIII-*Pst*I fragment of this extension demonstrated that it is expressed in forelimb blastema (data not shown). These overlapping cDNA clones establish an ORF of 454 amino acids with a predicted M_r of 51,302 (Fig. 1*b*). The initiation ATG identified in Fig. 1*b* is tentative because no upstream in-frame stop codon was present in the clone, but the surrounding sequence does largely conform with that anticipated for sites of translation initiation[19]. Sequence analysis indicates that this ORF encodes a full-sized RAR, with recognizable regions A–F. We designate this RAR as RARδ.

To determine whether the two newt RAR clones could function as RA-inducible *trans*-acting factors, cDNAs corresponding to the entire ORFs of the newt RARα and RARδ were assembled in the eukaryotic expression vector pSG5 to give vectors NvRARα and NvRARδ, respectively. These expression vectors were introduced into HeLa cells in transient transfection assays that were previously employed to characterize the mammalian RARs[16], and a reporter plasmid $(TRE3)_3$-tk-CAT was used to determine *trans*-activational efficiencies. This plasmid contained an RA responsive element that controlled the expression of the gene for chloramphenicol acetyl transferase (CAT), which could be readily monitored. Both NvRARδ and NvRARα showed reproducible RA-dependent activation of the $(TRE3)_3$-tk-CAT reporter, with activity detectable at RA concentrations as low as 10^{-9} M and maximal at an RA concentration of 10^{-7} M (see Fig. 2, top). The newt receptors were much less sensitive to retinol, showing little activity at a concentration of 10^{-7} M. The vector NvRARα was as effective at inducing reporter construct activity as its human counterpart (see Fig. 2, bottom), producing 14 to 20-fold increases in reporter construct expression at 10^{-7}–10^{-6} M RA. By contrast NvRAR-δ showed lower levels of *trans*-activation (~5-fold) when exposed to these levels of RA. The basis of this difference is the CAT activity detectable in the

absence of exogenous RA in the NvRARδ transfectants (see Fig. 2).

The newt RARδ is clearly distinct from the newt RARα. First, at the nucleotide level the two receptors are only 68% identical, which is virtually the same score that is obtained when the newt RARδ is compared with the human RARα (69%). Second, genomic Southern-blot analysis of homologous regions encoding the two receptors indicates that they arise from distinct genomic locations (data not shown). Third, the pattern of expression of the two receptors is different. In normal and regenerating limbs and tails, RARα is detected as two transcripts of ~6.3 and ~7.2 kb, whereas RARδ is abundantly expressed as a major transcript of ~6.7 kb (Fig. 3).

Comparison of amino-acid sequences indicates that RARδ is closely related to the recently described mammalian γ-receptor[16,17] in regions B–E and in immediately adjoining parts of regions A and F, but that the sequence of RARδ diverges markedly from that of RARγ at the N and C termini. Comparisons between the newt and human α-receptors (Table 1)

TABLE 1 Amino-acid sequence identity of newt and human RARs

	Overall	A	B	C	D	E	F
Newt RARα							
vs human RARα	91%	76%	96%	98%	83%	97%	77%
vs human RARβ	74%	(10%)	85%	95%	80%	91%	36%
vs human RARγ	69%	(19%)	81%	95%	65%	84%	(27%)
Newt RARδ							
vs human RARα	69%	(18%)	78%	95%	67%	85%	(16%)
vs human RARβ	73%	(15%)	85%	94%	67%	90%	(13%)
vs human RARγ	82%	39%	100%	98%	78%	96%	26%
Newt RARα vs RARδ	69%	(17%)	81%	94%	70%	86%	(23%)

Amino-acid sequence identity scores are given for comparisons of the α- and δ-newt RARs with each of the human RARs (α, β and γ)[12,15,17], and with each other. Listed are calculations for the full sequences and for pair-wise comparisons for each region A–F. Scores in parentheses indicate comparisons that did not identify shared motifs (that is, runs of identical or homologous residues). vs, versus.

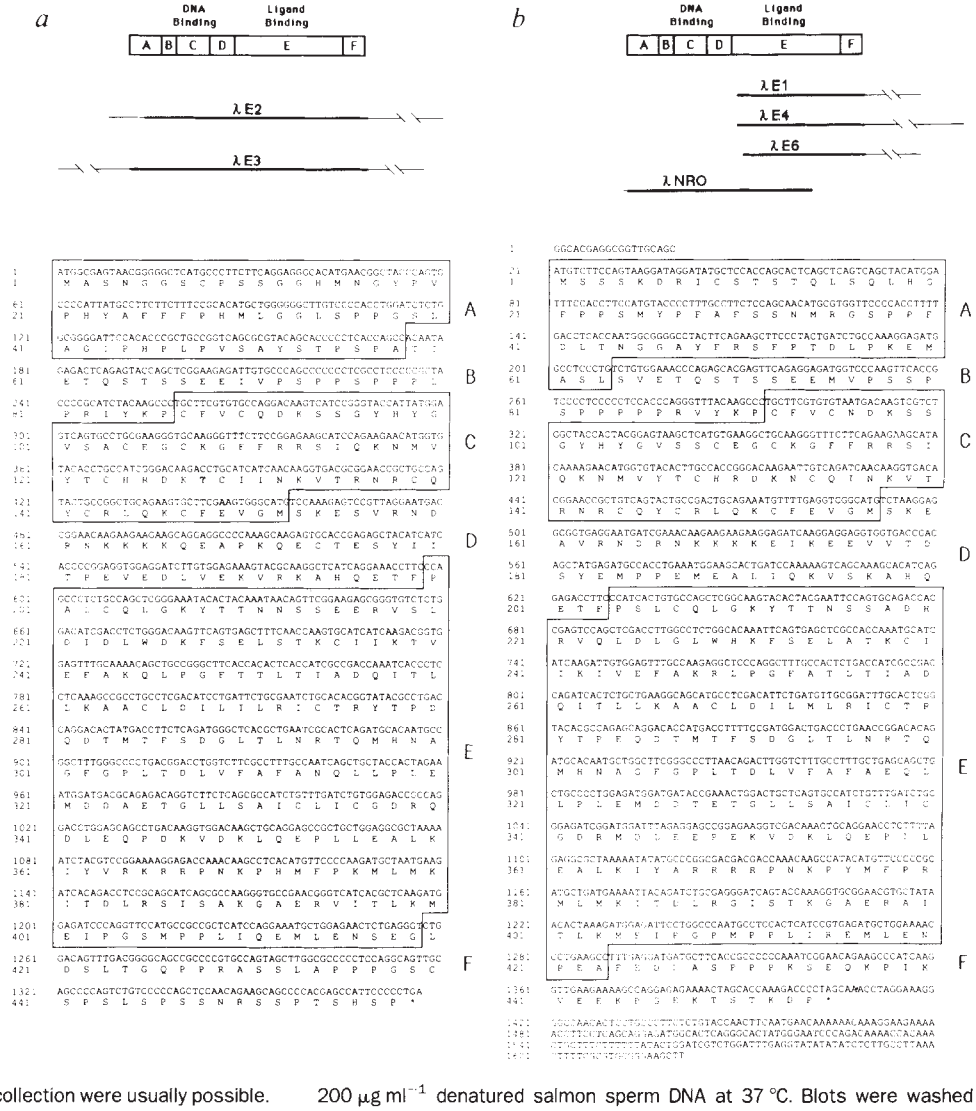


FIG. 1 Clone structure and nucleotide sequence data and predicted amino-acid sequence data for newt RARα (*a*) and RARδ (*b*). Single-letter amino-acid code is used. The cDNA's sketched at the top of the figure are aligned with a diagram illustrating the modular structure (regions A–F) of hormone nuclear receptors[26]. Region E is believed to bind to the hormone; region C is believed to bind to specific DNA sequences that are *cis*-acting hormone-responsive elements for targeted genes[26]. The N-terminal regions (A and B) are known to be important for promoter and cell-type specificity in transcriptional activation[27], but the C-terminal F region has not yet been shown to be functionally important. All six regions are indicated on the right side of the sequences and regions A, C and E are boxed. Nucleotides and amino acids are numbered on the left side of both sequences. The nucleotide sequences predicted by the cDNAs were not identical and several 'silent' polymorphisms were noted (for example a C→T substitution in the λE4 cDNA at position 1,098 of the RARδ gene sequence, which creates an *Xba*I site in this clone). *a*, Sequence of clone λE3. The λE2 sequence is as that of the λE3 clone from nucleotide 66, but an upstream termination codon in this cDNA predicts an N-terminal truncated receptor of 429 amino acids (compare with ref. 11).

METHODS. Healthy adult newts (Nv, *Notophthalmus viridescens*; Blades Biologicals, Edenbridge, Kent) were anaesthetized in a bath of 0.1% tricaine before surgery. Limbs were amputated in the middle of the upper forearm and the exposed bone and surrounding soft tissues were trimmed to produce a flat wound[28]. The newts were placed overnight in 0.5% sulfamerazine to recover. When the blastemas of the regenerating forelimbs reached the mid-bud stage (~16 days post-operatively), they were collected by amputation. Two additional rounds of tissue collection were usually possible. Tails were cut one-third to one-half the distance from their tips, and the amputated tissue was processed for normal tail RNA preparations[29]. From 10 μg poly(A)$^+$ RNA from forelimb blastema an oligo(dT)-primed cDNA library in the *Eco*RI site of λgt11 was provided (Clontech). A random-primed newt-tail cDNA library in the *Eco*RI site of λZAP (Stratagene) was prepared (Dr P. Savard). The forelimb blastema cDNA library was screened at low stringency with a 1,641-bp *Kpn*I–*Eco*RI fragment of the human RARα cDNA p63 (ref. 11). Hybridization was carried out over 2 days in 43% formamide, 5× SSC, 5× Denhardt's solution, 1% SDS, sodium phosphate (ph=6.5), 200 μg ml^{-1} denatured salmon sperm DNA at 37 °C. Blots were washed for 1 h at 37 °C in 2× SSC with 0.1% SDS, dried, and exposed to Kodak XAR-5 film at −70 °C with intensifying screens. Selected positive plaques were isolated and inserts were subcloned into the Bluescribe vector (Stratagene). Hybridization for the high stringency screening of the λZAP library was carried out at 42 °C in 50% formamide, and was followed by a 65 °C wash in 0.1× SSC, 0.1% SDS. Nucleotide sequences of the cDNA subclones were determined with the dideoxynucleotide method in the Sequenase DNA sequencing kit (US Biochemical Corp.).

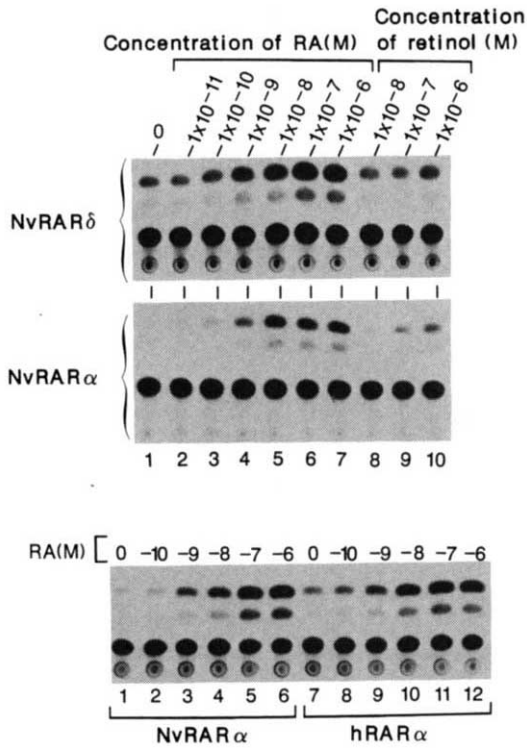


FIG. 2 Newt α- and δ-RARs function as RA-inducible *trans*-activating factors. HeLa cells were transfected with cDNA constructs of newt RARα (NvRARα), newt RARδ (NvRARδ) or human RARα (hRARα), together with the reporter plasmid $(TRE3)_3$-tk-CAT. After 24 h, the cells were fed with media that was free of retinoid (lane 0), or contained RA or retinol in the concentrations indicated in the lanes (1×10^{-11}–1×10^{-6} μ). The cells were harvested 48 h after transfection and levels of CAT activity determined.

METHODS. For the construction of the newt RARα expression vector (NvRARα), a 1.6-kb *Eco*RI–*Kpn*I fragment (pE3KR, from the clone λE3) was ligated into Bluescribe M13⁺ vector together with a 1.6-kb *Kpn*I–*Sph*I fragment (pE3Ka), also derived from λE3. From the resulting clone, an *Eco*RI–*Cho*I fragment containing the entire newt RARα ORF was transferred into the eukaryotic expression vector PTL1 (a pSG5-based vector containing an expanded polylinker). The extensive 1.4-kb 5′ untranslated region of the resulting construct was removed by inserting an *ECo*RI linker into a *Sna*BI site located 170 bp upstream of the first methionine codon of the ORF, discarding the resulting *Eco*RI fragment and re-ligating the vector. For the construction of the newt RARδ expression vector, NvRARδ, a 1.5-kb *Eco*RI fragment from the λE1 clone (pE1) containing the 3′ half of the NvRARδ ORF was ligated into the *Eco*RI site of pSG5. The ORF was completed by inserting a 640-bp *Sac*I fragment from the λZAP clone λNRO into the *Sac*I site of the pE1 fragment in PSG5 vector. In the transient expression studies, 2 μg NvRARδ, 1 μg NvRARα or 1 μg hRARα were co-transfected into HeLa cells (grown in 90-mm dishes) together with 2 μg reporter construct $(TRE3)_3$-tk-CAT (H. de Verneuil, D. Metzger and P. C., manuscript in preparation; see ref. 16), 2 μg β-galactosidase expression vector pCH110 (Pharmacia; added to normalize transfection efficiencies), and sufficient carrier DNA (BS M13⁺) to bring the final amount of DNA transfected to 20 μg per 90-mm dish. Transfection and retinoid treatment protocols were carried out as previously described[11,14,16], except that 50 OD_{420} units of β-galactosidase activity were used for the CAT assays shown for NvRARδ transfections, whereas 20 OD_{420} units were used for the NvRARα and the hRARα transfections.

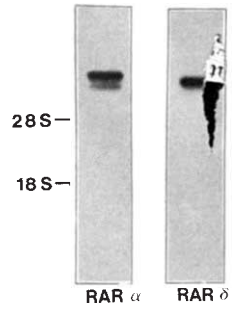


FIG. 3 Northern-blot analysis of the distribution of newt RARα and RARδ transcripts. Illustrated is a blot of poly(A)⁺ RNA from normal forelimb probed first for RARα (14 days exposure) and then for RARδ (9 h exposure). Similar transcription patterns were observed with forelimb blastema and normal and blastemal tail preparations. The positions of newt 28S and 18S rRNA are indicated on the left.

METHODS. Total RNA was extracted by homogenization in 4M guanidine thiocyanate[30] and then by centrifugation in phenol–chloroform, and poly(A)⁺ RNA was selected by one or two passages over an oligo(dT) column[31]. Samples of 3–8 μg mRNA were run on an agarose-formaldehyde gel and transferred to GeneScreen (NEN)[29]. A 1.6-kb NvRARα *Kpn*I–*Sph*I fragment (pE3Ka; includes ~0·6 kb of 3′ untranslated sequence) and the 939-bp NvRARδ *Eco*RI–*Hinc*III fragment (pE1c) were chosen as probes. Fragments were radiolabelled by random priming (BCL), and hybridizations were carried out overnight at 42 °C in 50% formamide, 1 M NaCl, 5× Denhardt's solution 2% SDS, 10% dextran sulfate, 50 mM Tris-Hcl (pH=7.5), 200 μg ml^{-1} denatured salmon sperm DNA. Blots were washed at 65 °C in 0.1× SSC, 2% SDS for 1 h and exposed to Hyperfilm (Amersham) with intensifying screens. Hybridized probe was stripped from blots before being used again in 96% formamide, 10 mM Tris-HCl (pH=7.5), 10 mM EDTA for 30 min at 60 °C.

and between the mouse and human γ-, β- and α-receptors[17] strongly suggest that if RARδ is the newt homologue of RARγ, then the sequence throughout regions A and F should be highly conserved. The presence of shared short runs of amino-acid sequence between RARδ and RARγ in regions A and F is insufficient to demonstrate homology, given that the mammalian α- and β-receptors share a Ser-Pro-Ser-X-Ser-Pro-Ser-Ser amino-acid motif in the F region. Although clear resolution of this issue depends on the successful identification of a newt RARγ, the structural similarity of RARδ and RARγ does raise the possibility that these receptors form a subclass of RARs that is distinguisable from the subclass that includes α- and β-receptors. In particular, the observation that the amino-acid sequence of the Cl region (which confers sequence specificity in DNA binding to target hormone-responsive elements[20,21]) is identical for RARα and RARβ, and for RARγ and RARδ, but differs by 2 out of 31 residues between RARα/β and RARδ/γ, indicates that δ- and γ-receptors could share a selectivity for natural hormone-response elements that is different from that of α- and β-receptors.

The presence of a functional RAR in the regenerating newt limb was recently reported[18]. The sequence of this receptor was based on two non-overlapping cDNA clones, isolated from a hindlimb blastema library, that were juxtaposed at an *Eco*RI site. The 3′ clone, λNB6, covers most of the E region, the F region and a considerable extent of 3′ untranslated sequence. In its translated region, the λNB6 cDNA is 98% identical at the nucleotide level with our RARδ clones and differs at only 2 out of 240 residues in predicted amino-acid sequence, indicating that it identifies the RARδ. The 5′ clone[18] λNB8, which covers a part of the A region, regions B–D and a part of region E, is only 75% identical in nucleotide sequence with our RARδ cDNA extension, λNRO. At the predicted amino-acid level, λNB8 is closer to the human RARβ (96% identity) than to the newt RARδ (79%) or the newt RARα (84%), and shows identity scores wth the human RARβ in regions B, C and D of 100%, 98% and 91%, respectively (compare with Table 1). We conclude first, that the composite receptor reported[18] is a chimaera, made up of RARδ and RARβ-like cDNA clones; and, second from *in situ* hybridization experients using part of the λNB8 clone[18], that the newt limb blastema expresses a RARβ-like RAR.

The multiple patterning effects produced by RA treatment of regenerating limbs (that is, proximal–distal and anterio–posterior axis re-specification as well as teratogenic-like deletions and complete inhibition of regeneration[22,23]) is consistent with the presence of at least three distinct RARs (α, δ and β-like) in limb blastemas. The challenge is to identify which receptors can control axis specification, and to determine whether a particular axis is governed by a unique RAR. Functional analysis of the receptors present in the blastema will be assisted by studies of their distribution and affinity for RA[24], and

examination of the effect of manipulating their expression in cultured blastemal cells that are subsequently introduced into regenerating limbs[25]. □

Received 11 August; accepted 8 September 1989.

1. Wallace, H. *Vertebrate Limb Regeneration* (Wiley, Chichester, 1981).
2. Niazi, I. A. & Saxena, S. *Folia Biol., Krakow* **26,** 3-11 (1978).
3. Maden, M. *Nature* **295,** 672-675 (1982).
4. Stocum, D. L. & Crawford, K. *Biochem. Cell Biol.* **65,** 750-761 (1987).
5. Brockes, J. P. *Neuron* **2,** 1285-1294 (1989).
6. Maden, M. *Devl. Biol.* **98,** 409-416 (1983).
7. Kim, W.-S. & Stocum, D. L. *Devl. Biol.* **114,** 170-179 (1986).
8. Maden, M., Keeble, S. & Cox, R. A. *Wilhelm Roux Arch. dev. Biol.* **194,** 228-235 (1985).
9. Tickle, C., Alberts, B., Wolpert, L. & Lee, J. *Nature* **296,** 564-565 (1982).
10. Summerbell, D. *J. Embryol. exp. Morph.* **78,** 269-289 (1983).
11. Petkovich, M., Brand, N. J., Krust, A. & Chambon, P. *Nature* **330,** 444-450 (1987).
12. Giguere, V., Ong. E. S., Segui, P. & Evans, R. M. *Nature* **330,** 624-629 (1987).
13. de The, H., Marchio, A., Tiollais, P. & Dejean, A. *Nature* **330,** 667-670 (1987).
14. Brand, N. J. *et al. Nature* **332,** 850-853 (1988).
15. Benbrook, D., Lernhardt, E. & Pfahl, M. *Nature* **333,** 669-672 (1988).
16. Zelent, A., Krust, A., Petkovich, M., Kastner, P. & Chambon, P. *Nature* **339,** 714.-717 (1989).
17. Krust, A., Kastner, P., Petkovich, M., Zelent, A. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **86,** 5310-5314 (1989).
18. Giguere, V., Ong, E. S., Evans, R. M. & Tabin, C. *J. Nature* **337,** 566-569 (1989).
19. Kozak, M. *Nucleic Acids Res.* **15,** 8125-8148 (1987).
20. Green, S., Kumar, V., Theulaz, I., Wahli, W. & Chambon, P. *EMBO J.* **7,** 3037-3044 (1988).
21. Mader, S., Kumar, V., de Verneuil, H. & Chambon, P. *Nature* **338,** 271-274 (1989).
22. Maden, M. *J. Embryol. exp. Morph.* **77,** 273-295 (1983).
23. Scadding, S. R. & Maden, M. *J. Embryol. exp. Morph.* **91,** 19-34 (1986).
24. Scadding, S. R. & Maden, M. *J. Embryol. exp. Morph.* **91,** 35-53 (1986).
25. Ferretti, P. & Brockes, J. P. *J. exp. Zool.* **247,** 77-91 (1988).
26. Krust, A. *et al. EMBO J.* **5,** 891-897 (1986).
27. Tora, L., Gronemeyer, H., Turcotte, B., Gaub, M.-P. & Chambon, P. *Nature* **333,** 185-188 (1988).
28. Kintner, C. R. & Brockes, J. P. *J. Embryol. exp. Morph.* **89,** 37-55 (1985).
29. Savard, P., Gates, P. B. & Brockes, J. P. *EMBO J.* **7,** 4275-4282 (1988).
30. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18,** 5294-5299 (1979).
31. Maniatis, T., Fritsch, E. & Sambrook, J. *Molecular Cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

ACKNOWLEDGEMENTS. We thank Tom Lufkin for the gift of pTL1, and Pierre Savard, Robin Brown and Geoff Scrace for their help and advice throughout this study. C.W.R. is a postdoctoral fellow of the National Multiple Sclerosis Society. M.P. was supported by a fellowship from FRMF.

Transactivation of the *Xenopus* rRNA gene promoter by its enhancer

Marietta Dunaway & Peter Dröge

Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA

A KEY question concerning the mechanism of transcriptional activation by enhancers is about the role of the DNA that connects the enhancer to the promoter[1-3]. The linking DNA will be important if a regulatory protein(s) binds to the enhancer and then tracks or slides along the DNA to the promoter, or if, on binding, the protein(s) alters the topological state of the DNA. By contrast, if the linking DNA loops out to allow the formation of a promoter–enhancer complex, or if the enhancer increases the local concentration of a transcription factor, co-linearity of the promoter and the enhancer will not be strictly required. In *Xenopus laevis*, the transcription of the ribosomal RNA genes is stimulated by an enhancer composed of repetitive sequences in the intergenic spacer regions[4,5]. These repetitive elements contain 60 or 81 base pairs, and their activity is relatively independent of their position and orientation[6-8]. When the enhancer and promoter sequences are each located on separate DNA molecules, however, the enhancer is no longer able to augment transcription[8]. We have now tested whether or not this apparent requirement for having the enhancer and promoter in *cis* can be overcome by keeping them in close proximity while locating them separately on different molecules. This was achieved by generating multiply intertwined, dimeric catenanes in which the enhancer and promoter were located in *trans* on different rings. By injecting these catenanes into frog oocytes and measuring the activity of the enhancers in a series of competition assays, we were able to demonstrate that such enhancers can augment transcription *in vivo*.

Dimeric catenanes have been used successfully to study how distant protein-binding sites interact during recombination[9,10]. We constructed a plasmid, pPD40E, that contained directly repeated sites for recombination by phage λ integrase, a rRNA minigene and an enhancer module of 10 repetitive elements (Fig. 1*a*)[5,8,11]. In this plasmid the enhancer is located ~3.7 kilobases (kb) from the promoter. To obtain multiply intertwined catenanes in which the enhancer and the promoter are located separately on different DNA molecules, we recombined pPD40E *in vitro* by using phage λ integrase, and the catenated material was separated from the unrecombined plasmid by electrophoresis (Fig. 1*b*)[12,13]. In plasmid pXlr401, the enhancer is in its normal position and orientation relative to the promoter (Fig. 1*c*).

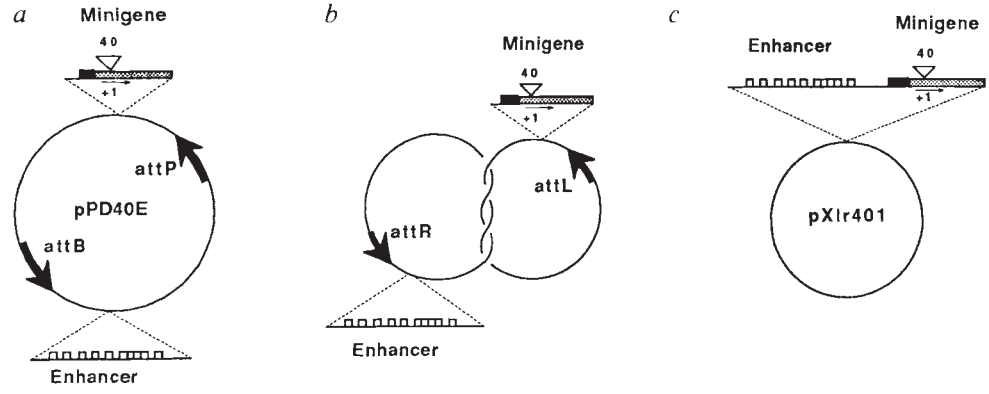


FIG. 1 Constructs used to test transactivation by the rRNA-gene enhancer. *a*, Plasmid substrate for site-specific recombination *in vitro* is pPD40E. The enhancer (open boxes) is near the *attB* site and the ψ40 minigene is near the *attP* site. The minigenes in pPD40E, pPDψ40, and pXlr401 contain an insert of 40 bp (open triangle) to distinguish their transcripts from the endogenous rRNA transcripts. The minigenes in pPDψ52 and pXlrψ52 contain 52-bp inserts so that transcripts from competing templates can also be distinguished. The black boxes represent the rRNA-gene promoter; the gene body (stippled box) is a fusion between the first 115 bp of the rRNA transcript with the final 200 bp of the 28S portion of the gene. The recombination sites are represented by black arrows. The distance between enhancer and promoter in pPD40E is ~3.7 kb. *b*, Recombination of pPD40E *in vitro* using phage λ integrase and *Escherichia coli* integration host factor was performed as previously described[11,12]. After recombination, the mixture of negatively supercoiled catenanes was electrophoresed in a 0.8% high-resolution agarose gel, which separates the recombined catenanes from residual unrecombined plasmid[22]. The catenanes were isolated from the gel and prepared for injection using standard methods. *c*, Plasmid pXlr401[8] contains ten 60- or 81-bp repeats in their normal position and orientation, and the insert is cloned into pUC19. METHODS. Construct pPD40E was prepared from pAB7.0d, pXlrψ40 and pXlr14F[5,8,11]. The vector pAB7.0d was modified to contain single restriction sites for *Eco*RI and *Kpn*I by eliminating one of the two *Eco*RI sites and by introducing a *Kpn*I site at one of the two *Nde*I sites. The inserts derived from pXlrψ40 (minigene) and pXlr14F (enhancer) had linkers added as needed and were inserted as *Eco*RI or *Kpn*I fragments, respectively. The related plasmids pPDψ40 and pPDψ52 contain only the minigenes ψ40 and ψ52, respectively, which were inserted into pAB7.0d at the same position as the minigene in pPD40E.

The principle of the competition assays is as follows. Frog oocytes are co-injected with two different plasmid DNAs[4,5]. Each of the plasmids carries a rRNA minigene that contains an insert of either 40 or 52 (base pairs bp) (Fig. 1). The injected templates compete for a pool of transcription factors. This results in the production of transcripts, which, because of their different-sized inserts, can be biochemically distinguished from each other and from endogeneous rRNA. This is achieved by using S1 probes specific for each of the two different transcripts that originate from the competing templates[14]. When one of the plasmids contains an enhancer, the activity of the enhancer can be expressed as the ratio of the transcripts originating from the enhancer-containing plasmids to those originating from the plasmids devoid of enhancers[5]. The competition assay thus provides an internal control for variations in the injection technique and for the recovery of RNA from the injected oocytes.

Figure 2*a* shows the results of one such assay in which the enhancer catenated to the promoter was used. For every experiment, we co-injected the two minigenes in equal amounts to provide a control for differences in the specific activities of the S1 probes (lanes 1*a* and 1*b*). In experiment 1, the probe specific activity for the transcript with the 52-bp insert (pXlrψ52) was about fourfold that for the 40-bp insert transcript (lanes 1*a* and 1*b*). Injection experiments in which the two minigenes were inserted into the plasmids pBR322, pUC19 and pAB7.0d indicate that neither vector sequences nor plasmid size affects the activity of the rRNA promoter in the competition assay (data now shown). A 70-fold transcriptional advantage was conferred on plasmids with enhancer sequences in their normal position and orientation over those that contained a minigene only (lanes 2*a* and 2*b*). The unrecombined plasmid pPD40E, in which the enhancer was ~3.7 kb away from the promoter, showed a sixfold greater transcriptional activity than a promoter alone (lanes 3*a* and 3*b*). In the purified catenanes, no unrecombined plasmid was detected (Fig. 3, lanes 4 and *C*). Co-injection of the purified enhancer–promoter catenanes with a plasmid containing only the promoter led to an eightfold higher transcriptional activity of the catenanes than of the normal plasmids (Fig. 2*a*, lanes 4*a* and 4*b*). Catenation is required for the effect because co-injection of an uncatenated enhancer with the two minigenes does not affect the transcriptional ratio between them (compare lanes 1*a* and 1*b* with lanes 5*a* and 5*b*).

To exclude the possibility that the activation observed in the catenated promoter–enhancer results from an undetectable amount of unrecombined pPD40E, we added unrecombined pPD40E to the mixture of plasmids that we used in the experiments shown in Fig. 2*a* (lanes 5*a* and 5*b*). The addition of pPD40E at ~5% the concentration of the other plasmids did not significantly change the ratio of the two transcripts to one another (compare lanes 5*a* and 5*b* with lanes 6*a* and 6*b*), yet the added pPD40E was detected on the Southern blot (Fig. 3, lane 6). It is improbable, therefore, that the transcriptional activation observed with the catenated enhancer was due to a small amount of undetected templates that had the enhancer in *cis* with the promoter. Although previous experiments have shown that there is no detectable difference in the transcriptional activity of supercoiled, nicked or relaxed templates injected into oocytes[15], it is also formally possible that the topology of catena-

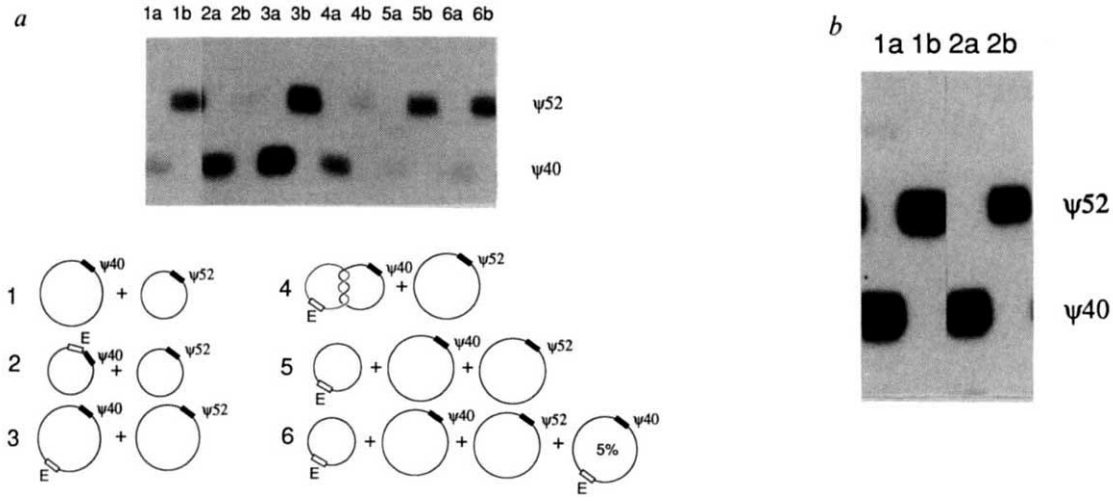


FIG. 2 Transcription of promoter–enhancer catenanes injected into *Xenopus* oocytes. *a*, As illustrated below the autoradiograph, injection mixtures are: (1), pPDψ40 + pXlrψ52 (2), pXlr401 + pXlrψ52 (3), pPD40E + pPDψ52 (4), pPD40E-catenanes + pPDψ52 (5), pPDψ40 + pPDψ52 + pXlr14F (6), reconstruction containing the mixture of plasmids used in lane 5 plus 5% added unrecombined pPD40E. *b*, In a separate experiment, catenanes of pPDψ40 were prepared and injected in competition with pPDψ52. The specific activities of the S1 probes in this assay were approximately equal. 1, pPDψ40 catenanes + pPDψ52; 2, unrecombined pPDψ40 + pPDψ52.

METHODS. DNAs were quantitated before injection by electrophoresis on agarose gels, ethidium bromide staining, and scanning of photographic negatives using a Hoefer densitometer. A solution (~20 nl) containing 500 pg of each competing template and 50 μg ml^{-1} α-amanitin was injected into each oocyte nucleus. Between 50 and 60 oocytes were injected with each sample. Transcripts were allowed to accumulate overnight, and the surviving oocytes (80% or more) were collected and pooled the next day. The oocytes were homogenized in a buffer containing 1% SDS, 100 mM NaCl, 50 mM Tris-HCl, pH 7.8, 20 mM EDTA and 0.1 mg ml^{-1} proteinase K. The mixture was incubated at 37 °C for 30 min, extracted in phenol and the nucleic acids precipitated with ethanol. The precipitate was resuspended in 10 μl sterile TE (10 mM Tris-HCl, 1 mM EDTA, pH 7.3) buffer per oocyte and used in all subsequent assays without further purification. Single-stranded, end-labelled probes that are specific for the transcripts of each of the two competing templates were prepared and hybridized in individual reactions with an equivalent of 5 oocytes from each sample[14]. After hydridization, the samples were treated with S1 nuclease, and the reaction was stopped by addition of excess EDTA. After precipitation with ethanol, the DNA was run on an 8% acrylamide–7M urea gel. The gel was dried and exposed to preflashed Kodak XAR film[23]. The final transcriptional ratios from competing templates were calculated as follows: Appropriate autoradiographic exposures of the gel were scanned to obtain uncorrected values for each transcription signal. Autoradiographs of the Southern blot in Fig. 3 were also scanned to determine the relative concentrations of competing templates. The transcription signals are then normalized to a template ratio of 1:1; if DNA is less than 5 ng, the transcriptional signal is proportional to the template concentration[24]. From this information, the relative specific activities of the S1 probes can be determined from the injection of the two promoters. The transcription results are then corrected for differences in probe specific activity to calculate the transcription ratios stated in the text. Lanes labelled as '*a*' were assayed with the ψ40 probe, and lanes labeled as '*b*', with the ψ52-specific probe.

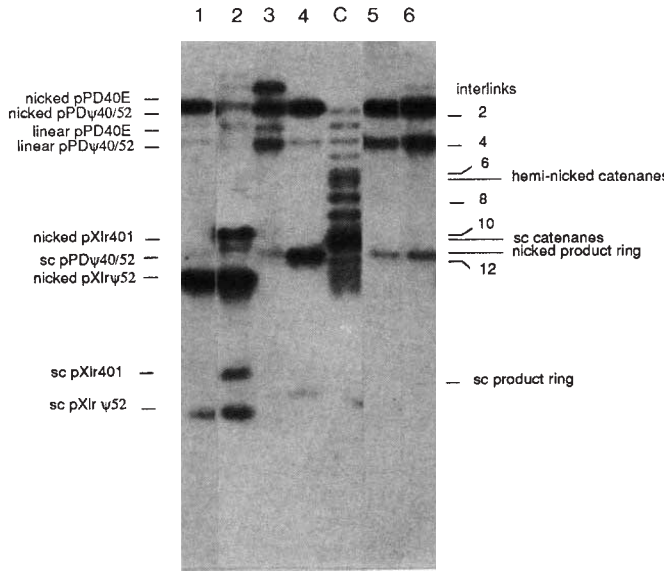


FIG. 3 Template analysis of the injected DNAs. Total nucleic acid from the equivalent of two injected oocytes was nicked by reaction with DNase I in the presence of 500 $\mu g\ ml^{-1}$ ethidium bromide. The samples were extracted in phenol, and an aliquot (1/3 oocyte) was electrophoresed as described[22]. The gel was blotted and probed with a fragment containing the ψ40 minigene. This probe will hybridize equally well with plasmids containing either the ψ40 or ψ52 minigene. The blot was exposed to film without using intensifying screens, and bands were quantitated by densitometry. Lane numbers correspond to those in Fig. 2*a*. Lane C, catenanes before injection; sc, negatively supercoiled.

tion *per se* could be responsible for the transcriptional activation of a promoter. To exclude this possibility, we prepared catenanes of a minigene without an enhancer by recombination of pPDψ40 *in vitro* and injected them in competition with unrecombined pPDψ52. Catenation did not change the ratio of the two transcripts (Fig. 2*b*).

Our results show that the enhancer for the rRNA minigene is not strictly required in *cis* with the promoter and can augment transcription when catenated to a promoter. We repeated the experiment four times and observed transactivation by the catenated enhancer in three experiments. In all cases, the level of enhancement was the same or higher than that observed with unrecombined pPD40E.

In the oocyte injection experiments, 18-20 hours elapsed between the injection of templates and the preparation of RNA. To assess both the extent of decatenation and the actual ratios of co-injected templates, the injected DNA was analysed by gel electrophoresis and Southern blotting. The catenanes contained between 2 and 12 interlinks before injection, and there was no detectable unrecombined plasmid (Fig. 3, lanes 4 and C). All of the interlinks were removed during an overnight incubation of the injected oocytes, however, so that only free-product rings were detected. A more detailed analysis of the kinetics of decatenation *in vivo* showed that this process is rapid and completed within 20-30 minutes after injection of catenanes (data not shown).

Because the *Xenopus* rRNA-gene enhancer can augment transcription in *trans*, enhancer models that invoke a protein-tracking mechanism or a DNA conformational change that is propagated through the linking DNA are ruled out. Models that are compatible with our results include the formation of a complex between the enhancer and promoter by looping, and the attachment of the enhancer to cell structures that mark active genes. But because the same transcription factor acts at both the promoter and the enhancer[25], the rRNA-gene enhancer may work by increasing the local concentration of this factor in the vicinity of a promoter[16]. If this is so, then the transcription factors could locate the promoter either by the looping out of the linking DNA or by three-dimensional diffusion.

The observation that the catenanes are completely resolved into product rings within 30 minutes of injection but are able to augment transcription, places further constraints on the viable models of enhancer action. Enhancement must occur rapidly after injection. It is possible that a complex between the promoter and enhancer is formed by protein-protein interaction, after which the two DNA rings can be unlinked without disrupting the enhancer-promoter complex. The rapid decatenation is also consistent with a requirement for the enhancer only during the initial activation of the gene. Experiments showing that enhancers can also function within the transcription unit further support the idea that the enhancer may only be required transiently[14]. The increasing evidence that RNA polymerases can transcribe through protein-bound DNA[17-20], however, suggests that the internal enhancer could still be part of a stable enhancer-promoter complex. Furthermore, the enhancer found in the intron of the immunoglobulin gene is required at least every cell cycle, and possibly continuously, for the expression of the gene[21]. Although our experiments do not distinguish these possibilities, we favour the formation of a stable promoter-enhancer complex. □

Received 4 May; accepted 11 September 1989.

1. Courey, A. J., Pion, S. E. & Wang, J. C. *Cell* **45,** 567-574 (1986).
2. Ptashne, M. *Nature* **322,** 697-701 (1986).
3. Echols, H. *Science* **233,** 1050-1056 (1986).
4. Moss, T. *Nature* **302,** 223-228 (1983).
5. Reeder, R. H., Roan, J. & Dunaway, M. *Cell* **35,** 449-456 (1983).
6. Boseley, P., Moss, T., Machler, M., Portmann, R. & Birnstiel, M. L. *Cell* **17,** 19-31 (1979).
7. Sollner-Webb, B. & Reeder, R. H. *Cell* **18,** 485-499 (1979).
8. Labhart, P. & Reeder, R. H. *Cell* **37,** 285-289 (1984).
9. Craigie, R. & Mizuuchi, K. *Cell* **45,** 793-800 (1986).
10. Kanaar, R., van de Putte, P. & Cozzarelli, N. R. *Cell* **58,** 147-159 (1989).
11. Wasserman, S. A., White, J. H. & Cozzarelli, N. R. *Nature* **334,** 448-450 (1988).
12. Mizuuchi, K., Gellert, M., Weisberg, R. A. & Nash, H. A. *J. molec. Biol.* **141,** 485-494 (1980).
13. Spengler, S. J., Stasiak, A. & Cozzarelli, N. R. *Cell* **42,** 325-334 (1985).
14. Labhart, P. & Reeder, R. H. *Nucleic Acids Res.* **13,** 8999-9009 (1985).
15. Harland, R., Weintraub, H. & McKnight, S. L. *Nature* **301,** 38-43 (1983).
16. Reeder, R. H. *Cell* **35,** 349-352 (1984).
17. Rhodes, D. *EMBO J.* **4,** 3473-3482 (1985).
18. Wolffe, A. P., Jordan, E. & Brown, D. D. *Cell* **44,** 381-389 (1986).
19. Solomon, M. J., Larsen, P. L. & Varshavsky, A. *Cell* **53,** 937-947 (1988).
20. Nacheva, G. A. *et al. Cell* **58,** 27-36 (1989).
21. Grosschedl, R. & Marx, M. *Cell* **55,** 645-654 (1988).
22. Sundin, O. & Varshavsky, A. *Cell* **25,** 659-669 (1981).
23. Laskey, R. A. & Mills, A. D. *FEBS Lett.* **82,** 314-318 (1977).
24. Windle, J. & Sollner-Webb, B. *Molec. cell. Biol.* **6,** 1228-1234 (1986).
25. Dunaway, M. *Genes Dev.* (in the press).

ACKNOWLEDGEMENTS. We thank N. R. Cozzarelli for suggestions throughout the work and for critical comments on the manuscript; B. Markoff, K. Scully and J. Spies for technical assistance, and C. Boles for comments on the manuscript. This work was supported by grants from the American Cancer Society and the NIH (M.D.). P.D. was supported by a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft.

RNA editing in wheat mitochondria results in the conservation of protein sequences

José M. Gualberto, Lorenzo Lamattina, Géraldine Bonnard, Jacques-Henry Weil & Jean-Michel Grienenberger*

Institut de Biologie Moléculaire des Plantes du CNRS, Université Louis Pasteur, 12, rue du Général Zimmer, 67084 Strasbourg cedex, France
* To whom correspondence should be addressed

RNA editing is a process that results in the production of a messenger RNA with nucleotide sequences that differ from those of the template DNA[1], and provides another mechanism for modulating gene expression. The phenomenon was initially described in the mitochondria of protozoa[2,3]. Here we report that RNA editing is also required for the correct expression of plant mitochondrial genes. It has previously been proposed that in plant mitochondria there is a departure from the universal genetic code[4], with CGG specifying tryptophan instead of arginine. This was because CGG codons are often found in plant mitochondrial genes at positions corresponding to those encoding conserved tryptophans in other organisms. We have now found, however, wheat mitochondrial gene sequences containing C residues that are edited to U residues in the corresponding mRNA sequences. In this way, CGG codons can be changed to UGG codons in the mRNA so that tryptophan may be encoded according to the universal genetic code. Furthermore, for each codon modification resulting from a C → U conversion that we studied, we found a corresponding change in the amino acid that was encoded. RNA editing in wheat mitochondria can thus maintain genetic information at the RNA level and as a result contribute to the conservation of mitochondrial protein sequences among plants.

Although it was proposed that in plant mitochondria CGG codes for tryptophan instead of arginine, no transfer RNA specific for tryptophan and recognizing CGG, or a gene for such a tRNA, has been found in plant mitochondria[5]. Moreover, the wheat mitochondrial genes *nad3*, *rps12* (ref. 6) and *cox3* (J.M.G., C. Domon, J.-H. W. and J.-M.G., manuscript in preparation) contain CGG codons corresponding to positions in their encoded amino-acid sequences that are conserved as either tryptophan or arginine residues. The recent characterization of a non-genomically encoded U at the junction of two exons in wheat *nad4* mRNA (L.L., J.-H.W., J.-M.G., manuscript submitted) showed that RNA editing can also exist in plant mitochondria. This encoding could result either from the addition of a U or from a C → U conversion. If such a C → U conversion affects a CGG codon, then the codon will be changed to UGG, which specifies tryptophan. To test whether this is occurring, we deduced the mRNA sequences corresponding to selected regions encompassing CGG codons of the wheat mitochondrial *nad3*, *rps12* and *cox3* genes. This was achieved by sequencing uncloned complementary DNAs that were primed by specific oligonucleotides, using total mitochondrial (mt) RNA as a template (Fig. 1). The CGG codons of *rps12* and *cox3*, which code for conserved arginines (positions indicated in Fig. 1) remain CGG codons at the RNA level. In *cox3*, however, the CGG codon located at a position corresponding to a conserved tryptophan is replaced by a UGG codon in the deduced mRNA sequence (Fig. 1). At a position in the *nad3* cDNA sequence that corresponds to a genomic CGG and a conserved tryptophan, either GCC or ACC is found. This ambiguity could not be resolved by using different RNA preparations as templates. The *nad3* mRNA seems, therefore, to exist as two distinct species, which have either CGG or UGG codons. This could be due to the presence of both unmodified and modified transcripts. We used the same approach to determine the cDNA sequence of regions of the wheat mitochondrial *cox1* (ref. 7), *cox2* (ref. 8) and *cob* (ref. 9) genes where CGG codons are present (Fig. 2). The results show that the three CGG codons of *cox2* and the two CGG codons of *cob* (all located at positions corresponding to conserved tryptophan residues) are all converted to UGG codons in the mRNAs. The only CGG of *cox1*, which corresponds to a non-conserved position, is unmodified (data not shown). It is not

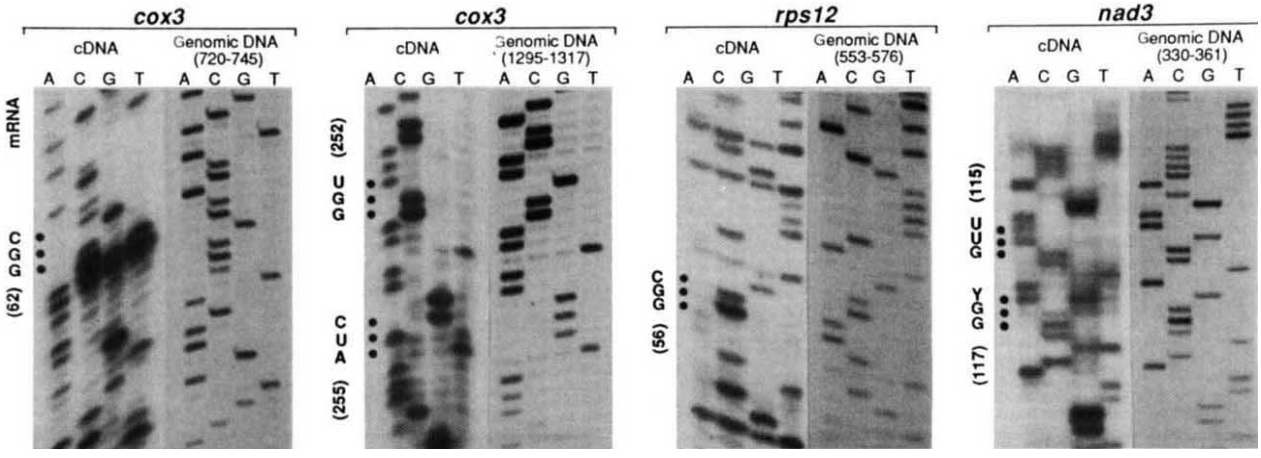


FIG. 1 Comparison of the cDNA and genomic DNA sequences in selected regions of *cox3*, *rps12* and *nad3*. For each gene, the right panel shows part of the cloned mt gene sequence (with the coordinates indicated in parentheses) and the left panel shows the corresponding cDNA sequence. The dots indicate the nucleotides of the mRNA codons deduced from the cDNA sequence, with codon positions given in parentheses (referred to in Table 1 and Fig. 3). For cDNA sequence, 5′-end ^{32}P-labelled synthetic oligonucleotides were hybridized with 70 μg of total wheat mtRNA in 14 μl of 250 mM KCl, 10 mM Tris-HCl, pH 8.3. Sequencing by the dideoxy chain-termination method using AMV reverse transcriptase was performed at 50°C for 60 min in 10 μl of a solution containing 3 μl of the primer-template RNA mixture, 8 mM $MgCl_2$, 0.5 mM of each dNTP, 4 mM DTT, 50 μg/ml actinomycin D and 0.2 mM of either ddATP, ddCTP, ddGTP or ddTTP. Oligonucleotides used as primers are complementary to the published gene sequences between the following positions: For *cox3* 762-786 and 1,332-1,356 (J.M.G., C. Domon, J.-H.W. and J.-M.G., submitted); for *nad3* 397–421 (ref. 6); for *rps12* 599–623 (ref. 6). Sequence of uncloned genomic mt DNA was obtained by the triple primer method as described[15], except that *Taq* DNA polymerase (Biolabs) was used for DNA amplification and T7 DNA polymerase (Pharmacia) was used for sequencing according to the manufacturer's instructions.

possible, therefore, to predict whether a CGG in a wheat mitochondrial gene will code for arginine or tryptophan, as this is dependent on whether RNA editing takes place. Our results show that as a result of RNA editing in wheat mitochondria, the universal genetic code can be used to specify tryptophan, without the requirement of divergence from the universal code postulated by Fox and Leaver[4].

By analysing cDNA sequences, we found differences between the deduced mRNA sequences and their corresponding gene sequences other than the ones involving CGG codons. We found 17 differences by comparing ~1,500 deduced mRNA nucleotides with the corresponding genomic sequences (Table 1). All these differences are C→U conversions, which occur at the first or second base of the codon and lead to a corresponding change of the amino acid encoded. But no modification was found in the 170 and 150 nucleotides sequenced in non-coding regions upstream of *cox3* and *cob*, respectively (data not shown), which indicates that the C→U editing acts to modulate the coding part of mRNA. In agreement with this, we found that the amino-acid sequences deduced from the edited mRNAs have more homology with those of the mitochondrial proteins of other species than do the amino-acid sequences deduced from the corresponding genes (Fig. 3). In some plants, the mitochondrial genome encodes amino acids in a particular protein that are the

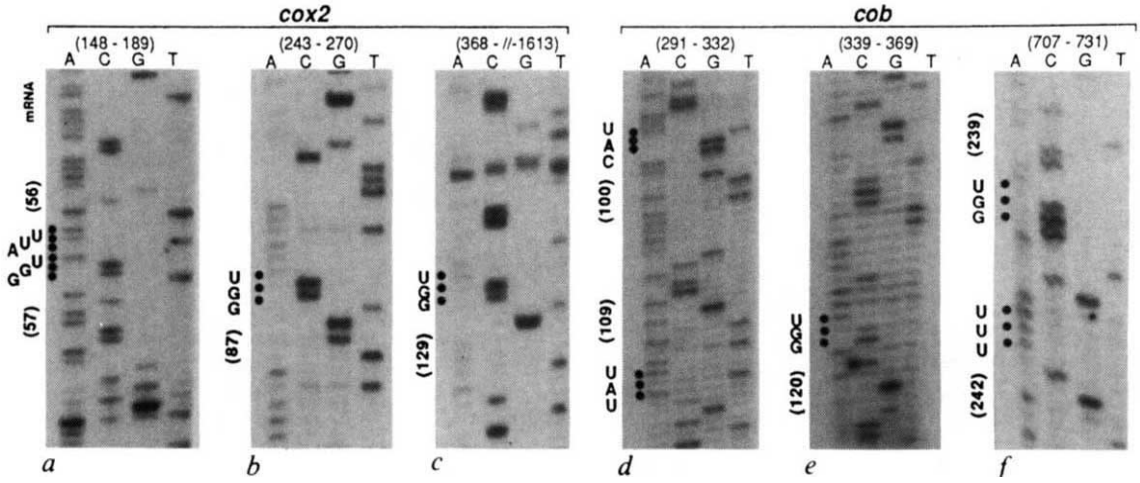


FIG. 2 Sequence of selected cDNA regions of *cox2* and *cob*. Sequences of the corresponding genomic DNA[8,9] are not shown, but their coordinates are given in parentheses. Codon positions are indicated as in Fig. 1. Sequence by the dideoxy chain-termination method using total mt RNA as template was as described in Fig. 1. Oligonucleotides used as primers are complementary to the published gene sequences between the following positions: For *cox2* (gels *a* and *b*), 292–316 (ref. 8); gel *c*, 1,638–1,662 (ref. 8); the latter primer allowed to sequence across the exons junction. For *cob* (gels *d* and *e*), 391–415 (ref. 9); gel *f*, 748–772 (ref. 9). For *cox1*, sequence not shown (1,474–1,498)[7].

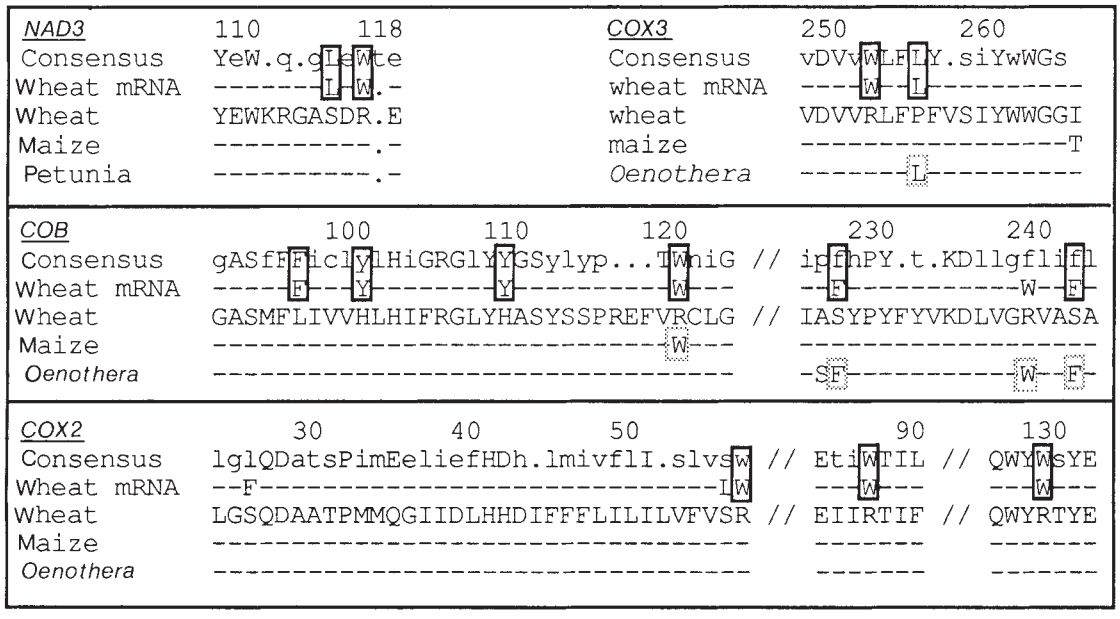


FIG. 3 Increases of mt protein sequence conservation due to C→U editing. Selected parts of the amino-acid sequence of several wheat mt proteins (deducted from the gene sequences using the universal genetic code) are compared with (1) the corresponding amino-acid sequence of mt proteins (deduced from the gene sequences) from two other plants (one monocot and one dicot), (2) the amino-acid sequence deduced from the wheat mRNA sequence, (3) the consensus amino-acid sequence based on mt protein sequences from non-plant species (found in GenEMBL Data Bank). Dots in the consensus sequence indicate non-conserved positions, lower case letters indicate partial conservation, capitals indicate complete conservation. A dash indicates that the amino acid is identical to that deduced from the wheat genomic sequence. Solid line boxes indicate amino-acid homology with the consensus sequence, when editing of the wheat mRNA is considered. Dotted line boxes indicate amino-acid homology with mt proteins of other plants, when editing of the wheat mRNA is considered.

TABLE 1 Differences found in several wheat mt genes between the codons in the genomic sequences and the codons in the mRNAs (as deduced from the cDNA sequences)

Gene		mRNA	Localization of modified codons
CGG	—	UGG	*nad3* (117): *cox2* (57, 87, 129)
(R	→	W)	*cox3* (252); *cob* (120, 239)
CAC	—	UAC	*cob* (100)
CAT	—	UAU	*cob* (109)
(H	→	Y)	
TCA	—	UUA	*cox2* (56)
TCG	—	UUG	*rps12* (24); *nad3* (115)
(S	→	L)	
TCT	—	UUU	*cob* (227, 242); *cox2* (26)
(S	→	F)	
CTC	—	UUC	*cob* (96)
(L	→	F)	
CCG	—	CUG	*nad4* (316)
CCA	—	CUA	*cox3* (255)
(P	→	L)	

The positions where a C→U editing has been observed are underlined and the amino acids specified by the corresponding codons are indicated using the one-letter code.

same as the corresponding amino acids specified after editing in wheat mitochondria (Fig. 3). Furthermore, in these cases the non-edited genomic codons are the same as the corresponding edited nongenomic codons. This indicates that RNA editing has a role in the conservation of mitochondrial protein sequences during evolution.

The C→U conversion detected in wheat mitochondria has so far not been studied in other plants, but it is likely to occur, given the conservation of plant mitochondrial gene sequences[10]. In fact, the sequences of two cDNA clones of the *Oenothera cox3* gene show differences from the corresponding genomic sequences that can be explained by C→U conversions[11].

Plant mtDNAs can comprise several subgenomic molecules generated by homologous recombination[12]. It is possible that the U-containing RNAs were transcribed from T-containing DNA molecules not present in the genomic libraries used for the sequencing studies. For the edited *cox3 and nad4* transcripts, however, the sequencing of uncloned genomic DNA amplified by the polymerase chain reaction did not yield results in support of this idea (data not shown). Because hybridization of *cox3* and *nad4* probes with total wheat DNA only detected fragments corresponding to the mtDNA restriction map (not shown), the existence of T-containing DNAs in the nucleus could also be ruled out. Sub-stoichiometric DNA molecules in plant mitochondria[13] could comprise T-containing DNAs, but they would have to be selectively transcribed, which is unlikely.

If one excludes, therefore, the possibility that undetected genomic DNA templates code for wheat mtRNAs, the presence of non-genomically encoded U residues at specific positions in mRNAs can be explained either by a misincorporation of U residues by the mtRNA polymerase, or by a C→U conversion, which possibly involves a deamination step. This could be related to the C→U conversion that is required for the expression of the nuclear gene encoding apolipoprotein B (ref. 14). In any case, as RNA editing occurs in plant mitochondria, the sequence of mitochondrial proteins cannot be deduced solely from the gene sequence and should be checked by sequencing the corresponding cDNA. Because RNA editing is important in the modulation of gene expression, the mechanisms that it involves and the stage of mRNA synthesis or maturation at which it occurs need to be determined. □

Received 21 August; accepted 11 September 1989.

1. Simpson, L. & Shaw, J. *Cell* **57**, 355–366 (1989).
2. Benne, R. *et al. Cell* **46**, 819–826 (1986).
3. Shaw, J. M., Feagin, J. E., Stuart, K. & Simpson, L. *Cell* **53**, 401–411 (1988).
4. Fox, T. D. & Leaver, C. J. *Cell* **26**, 315–323 (1981).
5. Maréchal, L., Guillemaut, P., Grienenberger, J. M., Jeannin, G. & Weil, J. H. *Nucleic Acids Res.* **13**, 4411–4416 (1985).
6. Gualberto, J. M., Wintz, H., Weil, J. H. & Grienenberger, J. M. *Molec. gen. Genet.* **215**, 118–127 (1988).
7. Bonen, L., Boer, P. H., McIntosh, J. E. & Gray, M. W. *Nucleic Acids Res.* **15**, 6734 (1987).
8. Bonen, L., Boer, P. H. & Gray, M. W. *EMBO J.* **3**, 2531–2536 (1984).
9. Boer, P. H., McIntosh, J. E., Gray, M. W. & Bonen, L. *Nucleic Acids Res.* **13**, 2281–2292 (1985).
10. Palmer, J. D. & Herbon, L. A. *J. molec. Evol.* **28**, 87–97 (1988).
11. Hiesel, R., Schobel, W., Schuster, W. & Brennicke, A. *EMBO J.* **6**, 29–34 (1987).
12. Newton, K. *J. A. Rev. Pl. Physiol. Pl. molec. Biol.* **39**, 503–532 (1988).
13. Small, I. D., Isaac, P. G. & Leaver, C. J. *EMBO J.* **6**, 865–869 (1987).
14. Powell, L. N. *et al. Cell* **50**, 831–840 (1987).
15. Wrischnik, L. A. *et al. Nucleic Acids Res.* **15**, 529–542 (1987).

ACKNOWLEDGEMENTS. J.M.G. is supported by a fellowship from the French Ministère des Affaires Etrangères and by the Ministry of Education of Portugal. L.L. is supported by a fellowship from Conicet (Argentina).

RNA editing in plant mitochondria

Patrick S. Covello & Michael W. Gray*

Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

A BASIC principle of molecular biology is that the primary sequence of RNA faithfully reflects the primary sequence of the DNA from which it is transcribed. This concept has been challenged recently by the discovery of RNA editing[1], broadly defined as any process that changes the nucleotide sequence of an RNA molecule from that of the DNA template encoding it. Examples of RNA editing (see ref. 2 for review) include the insertion and deletion of uridine residues in mitochondrial messenger RNAs in kinetoplastid protozoa[3], the conversion of a cytidine to uridine in mammalian apolipoprotein-B mRNA[4–7], and the appearance of two non-templated guanosine residues in a paramyxovirus transcript[8,9]. In these cases, RNA editing either re-tailors a non-functional transcript, producing a translatable mRNA[10,11], or modifies an already functional mRNA so that it generates a protein of altered amino-acid sequence[4–9]. Here we report an editing phenomenon that involves the conversion of cytidine to uridine at multiple positions in the mRNA for subunit II of cytochrome *c* oxidase in wheat mitochondria. Such RNA editing provides an explanation for apparent coding anomalies in plant mitochondria.

The alignment of the deduced amino-acid sequences of proteins encoded by mitochondrial (mt) DNA, such as that shown for cytochrome *c* oxidase (COXII) in Fig. 1, reveals several amino-acid positions that are invariant in non-plant species but seem to have undergone radical substitutions in one or more plant species. These substitutions include that of arginine for tryptophan at several positions (solid squares, Fig. 1), which originally prompted the suggestion that plant mitochondria use CGG to code for tryptophan rather than for arginine[12]. This is the only proposed deviation from the universal genetic code in plant mitochondria[13]. When, however, comparisons of other mitochondrial proteins are made, CGG codons in plant sequences align at positions specifying arginine residues in non-plant sequences[14,15], which implies that if CGG does code for tryptophan in plant mitochondria, it does not do so consistently. So far, no plant proteins encoded by mtDNA have been directly sequenced, so the issue of which amino acid is specified by CGG in plant mitochondria is unresolved. Furthermore, two separate transfer RNAs with different anticodons would be required if CGG as well as UGG were to specify tryptophan. As yet, only a single $tRNA^{Trp}$ that decodes UGG has been found in plant mitochondria[16,17].

It is also peculiar that although mitochondrial protein genes are highly conserved in the coding sequence among related plants, the few differences that are seen are expected to lead

*To whom correspondence should be addressed.

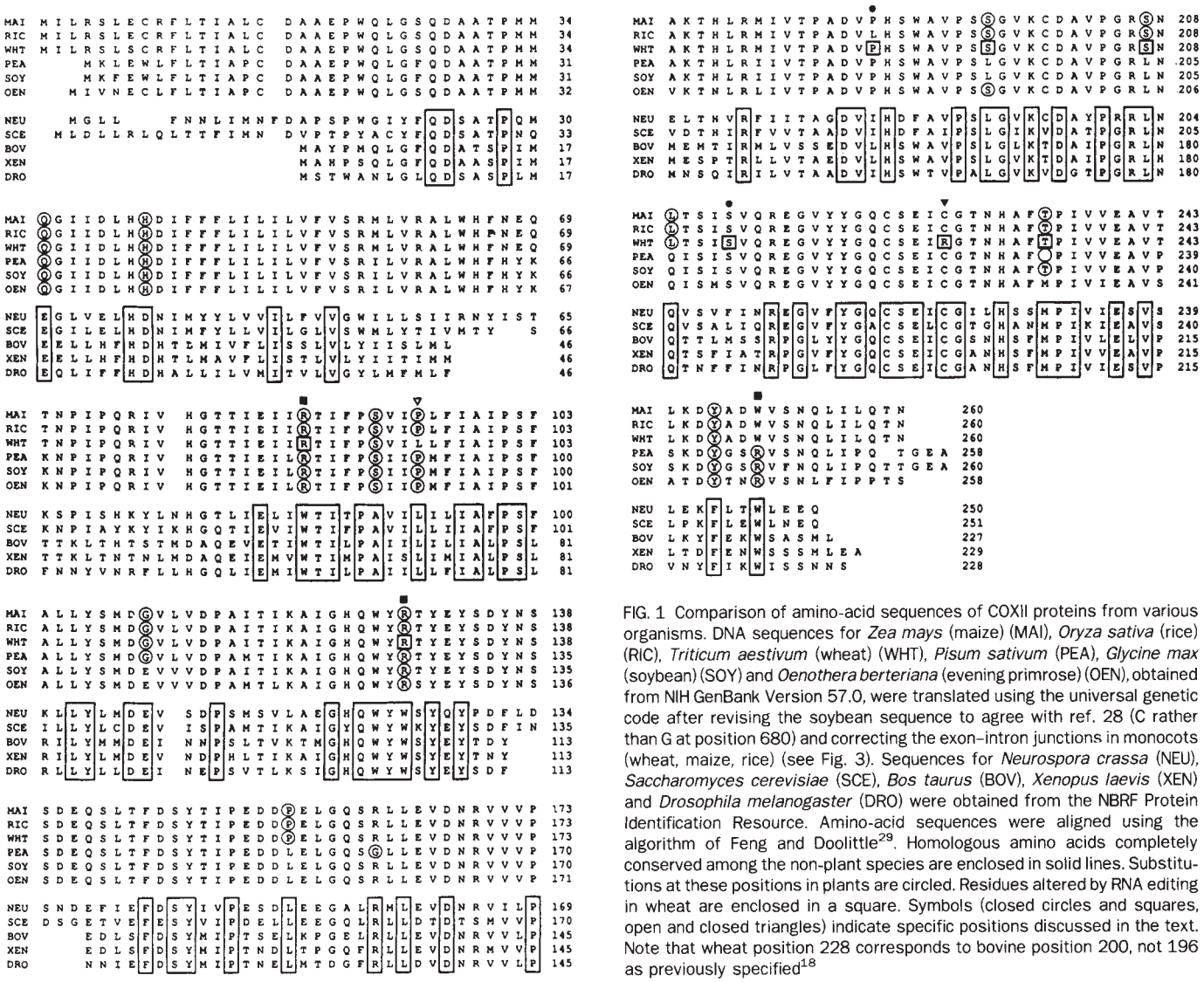


FIG. 1 Comparison of amino-acid sequences of COXII proteins from various organisms. DNA sequences for *Zea mays* (maize) (MAI), *Oryza sativa* (rice) (RIC), *Triticum aestivum* (wheat) (WHT), *Pisum sativum* (PEA), *Glycine max* (soybean) (SOY) and *Oenothera berteriana* (evening primrose) (OEN), obtained from NIH GenBank Version 57.0, were translated using the universal genetic code after revising the soybean sequence to agree with ref. 28 (C rather than G at position 680) and correcting the exon–intron junctions in monocots (wheat, maize, rice) (see Fig. 3). Sequences for *Neurospora crassa* (NEU), *Saccharomyces cerevisiae* (SCE), *Bos taurus* (BOV), *Xenopus laevis* (XEN) and *Drosophila melanogaster* (DRO) were obtained from the NBRF Protein Identification Resource. Amino-acid sequences were aligned using the algorithm of Feng and Doolittle[29]. Homologous amino acids completely conserved among the non-plant species are enclosed in solid lines. Substitutions at these positions in plants are circled. Residues altered by RNA editing in wheat are enclosed in a square. Symbols (closed circles and squares, open and closed triangles) indicate specific positions discussed in the text. Note that wheat position 228 corresponds to bovine position 200, not 196 as previously specified[18]

often to radical changes at the positions of otherwise highly conserved amino-acid residues. For example, the maize[12] and wheat[18] COXII-coding sequences differ at only nine positions in their 780-nucleotide coding region. Five of these substitutions are expected to result in three radical amino-acid changes in the wheat COXII sequence compared with that of the maize protein[18], which are (in the direction maize to wheat) Glu (GAA) to Ser (TCA) at position 7, Pro (CCA) to Leu (CTT) at position 95 (open triangle, Fig. 1) and Cys (TGT) to Arg (CGT) at position 228 (closed triangle, Fig. 1). The last difference is particularly puzzling because Cys 228 in maize COXII (and its equivalent in other COXII proteins) is thought to bind a catalytically essential copper atom.[19] Of all reported COXII sequences, only the wheat sequence apparently lacks cysteine at this particular position.

These irregularities led us to consider the possibility that in plant mitochondria, alterations in mRNA structure might act to maintain similarity at the protein level, despite sequence divergence at the gene level. A striking proportion (35 out of 56, or 63%) of the minimal changes that are required to convert anomalous codons in plant mitochondrial COXII mRNAs to codons that specify the consensus amino-acid residue in non-plant sequences are C→U conversions (Table 1). In fact, such C→U changes could resolve all of the coding anomalies in plant mitochondria mentioned above. In particular, selective conversion of CGG to UGG in mRNA would restore the expected (universal) codon at positions in plant mitochondrial proteins that in non-plants specify conserved tryptophan residues. Also, a C→U change in codon 228 (CGU, arginine) in the wheat COXII mRNA would switch this codon to UGU, permitting the incorporation of cysteine rather than arginine at position 228, so that the wheat COXII protein would now conform with all other COXII proteins at this site (solid triangle, Fig. 1).

The first indication that C→U RNA editing might actually occur in plant mitochondria came from the characterization of primary transcripts from wheat mitochondria (P.S.C. and M.W.G., manuscript in preparation). Direct sequence analysis of a 1.3-kilobase (kb) transcript, identified as COXII mRNA, showed that position −101 from the translation initiation site (within a 5′-untranslated leader of ~170-nucleotides) is occupied by U in the mRNA but by C in the previously determined sequence of the single-copy *coxII* gene[18] (Figs 2*a* and 3). We have observed the same difference between the gene and the mRNA sequence at the same position in a 0.8-kb transcript (corresponding to the mRNA of *orf25*, a gene of unknown function sequenced by L. Bonen; personal communication) that shares an identical 5′-terminal sequence of at least 85 nucleotides with COXII mRNA (data not shown).

We therefore examined the coding region in COXII mRNA by reverse transcriptase sequencing[20]. Figures 2*b*, *c* and Fig. 3

indicate the regions of RNA sequenced and draw attention to the nine differences found between the corresponding RNA and DNA sequences. These were all C in the DNA sense strand but U in the RNA. In the portions of COXII mRNA that were sequenced, the differences included all of the C→U changes indicated from amino-acid sequence comparison (Fig. 1 and Table 1), that is, codons 87 (Arg to Trp), 129 (Arg to Trp), 196 (Ser to Leu), 207 (Ser to Leu), 228 (Arg to Cys) and 235 (Thr to Met) (Fig. 3). In addition, codons 188 and 213 were both changed to specify Leu from specifying Pro and Ser, respectively (Fig. 3), consistent with the hydrophobic nature of the residues at the homologous positions in non-plant sequences (closed circles, Fig. 1).

Our observations provide evidence of a remarkable degree of C→U editing in wheat mitochondrial COXII mRNA. Such editing has the effect of altering the predicted COXII amino-acid sequence towards that of the non-plant consensus sequence at highly conserved positions, and explains the apparent coding anomalies in plant mitochondria. For instance, pending confirmation by direct COXII protein sequencing, wheat amino-acid residue 228 (corresponding to Cys 200 in bovine COXII[21]) can be reinstated as a universally conserved cysteine residue,

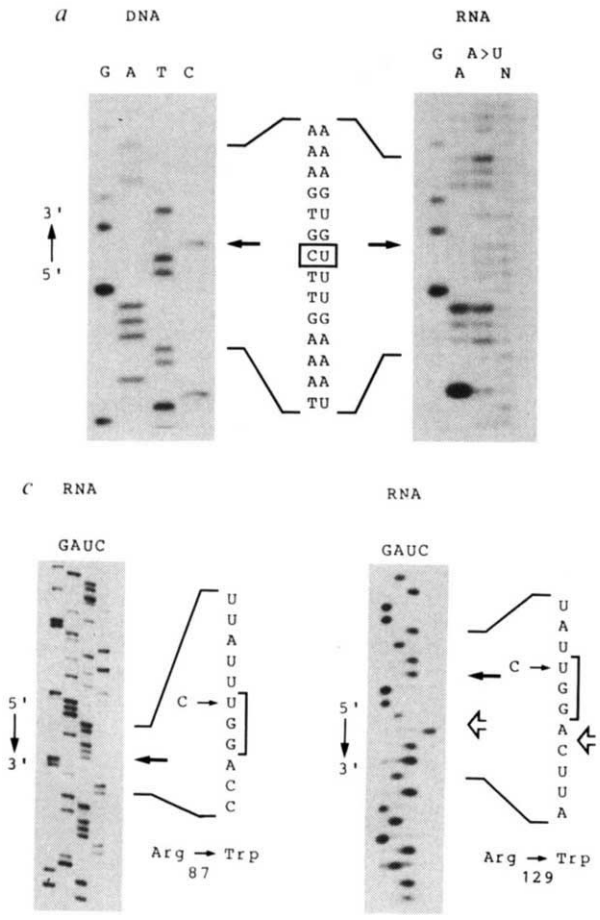
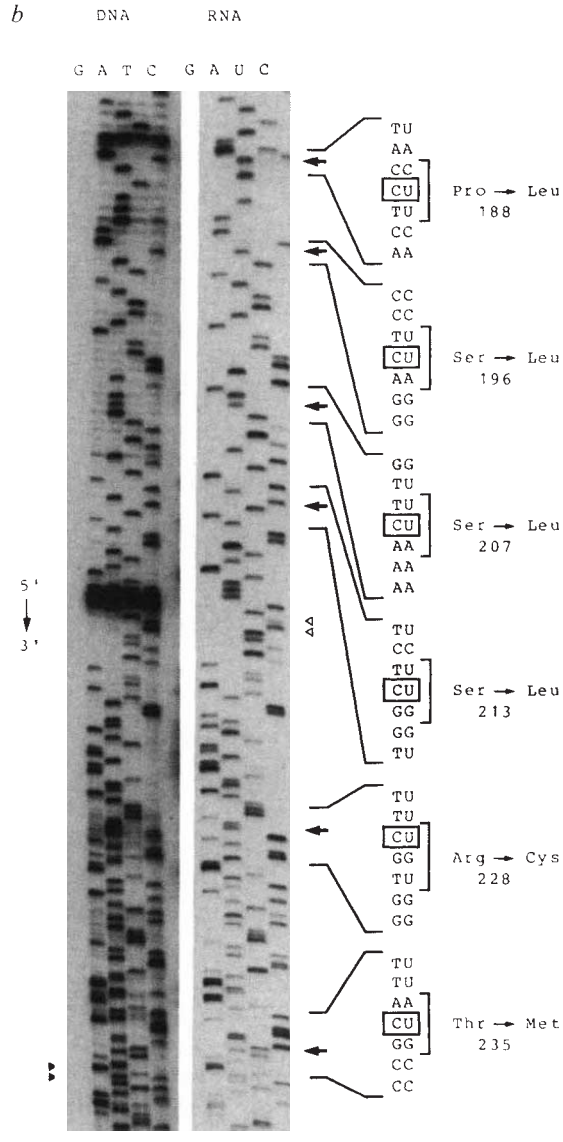


FIG. 2 Comparison of autoradiographs of DNA- and RNA-sequencing gels of the gene and mRNA for wheat mitochondrial COXII. *a*, 5′-Untranslated leader region showing C and U at homologous positions (arrows) in the DNA (RNA-like strand) and mRNA, respectively. The autoradiograph of the DNA sequencing gel was rotated 180° before being photographed to show an RNA-like sequence directly. *b*, Dideoxy sequence of DNA and RNA at the 3′ end of the COXII reading frame (exon 2), showing C→U editing of RNA (arrows). Autoradiographs have been positioned to display the complementary (RNA-like) sequence directly. Sequencing from DNA and mRNA templates was primed with the same oligodeoxyribonucleotide (primer 1, Fig. 3). Open triangles indicate ambiguous positions in the RNA sequencing gel, whereas the closed triangles show an area of compression in the DNA sequencing gel. The published DNA sequence[18] is shown for the latter region. *c*, Reverse transcriptase sequence of COXII mRNA across the exon 1–exon 2 boundary (open arrow) and through putative tryptophan codons (CGG) in exon 1 (solid arrows), using primer 2 (Fig. 3). Autoradiographs have been positioned to display the complementary (RNA-like) sequence.

METHODS. Primary transcripts (RNA species having pppN- or ppN-5′-termini) were capped[30] *in vitro* by incubation of NaCl-insoluble wheat embryo mitochondrial RNA[18,31] with [α-^{32}P]GTP and guanylyltransferase. A purified 1.3 kb band (isolated by PAGE) was sequenced by the enzymatic degradation procedure (see ref. 32) using RNases T1 (G), U2 (A), Phy M (A > U) and M1[33] (does not cut after C; data not shown). N, alkali. RNA sequence in the coding region was obtained by the dideoxy chain-termination method of reverse



transcriptase sequencing (ref. 17, procedure 2) using the primers 1 and 2 (sequence shown in lower-case letters in Fig. 3); these were extended in the direction indicated by the arrows in Fig. 3. To confirm the COXII gene sequence, M13 clones of wheat mtDNA, originally sequenced[18] using the Klenow fragment of *Escherichia coli* DNA polymerase I, were re-sequenced using the Sequenase system[34] (US Biochemicals). Note that in *a*, different varieties of wheat (Thatcher and Katepwa) were used for isolation of the COXII gene and of capped COXII mRNA, respectively. Therefore the difference between DNA and RNA observed in the 5′-untranslated region (position −101 in the COXII mRNA; Fig. 3) could be a manifestation of a strain-specific sequence difference at this position, rather than C→U RNA editing. For reverse transcriptase sequencing of the COXII mRNA coding region, however, three different mitochondrial RNA preparations (from Thatcher, Katepwa and Chinook wheat) were used as template with primer 1(*b*), whereas two preparations (Thatcher and Katepwa) were used with primer 2(*c*). Identical results were obtained with the different RNA preparations in each case.

FIG. 3. Comparison of gene and partial mRNA sequence for wheat COXII. The complete DNA sequence (ref. 18, and unpublished observations) is shown and below it, the amino-acid sequence inferred using the universal genetic code. Unambiguous RNA sequence is underlined: broken underline, direct sequencing (beginning near the 5′ end of the capped transcript); solid underline, reverse transcriptase sequencing. Differences from the corresponding DNA sequence and in the amino-acid sequence deduced from the mRNA sequence are shown above the DNA sequence and are boxed. RNA sequence in the coding region was obtained by the dideoxy chain-termination method of reverse transcriptase sequencing[20] using the primers 1 and 2 (sequence in lower case letters; Regional DNA Synthesis Laboratory, University of Calgary). Asterisks indicate ambiguous nucleotide positions in the RNA sequence. The vertical arrow indicates the position of insertion of the single intron in the gene, precisely defined by the RNA sequence spanning the exon 1–exon 2 boundary (Fig. 2*c*). From this information, the DNA sequence at the exon–intron boundaries must be CGGAGTGCG . . . CAACCTTAT (intron sequence underlined); that is, the boundaries are two nucleotides upstream of those suggested previously for the maize[12], wheat[18] and rice[35] COXII genes. This shift changes the predicted amino-acid residue at position 130 of the COXII protein from Ser (AGT) to Thr (ACT) (Fig. 3). The downstream triangle indicates a residue previously reported to be A (ref. 18).

TABLE 1 Possible sites of RNA editing in plant COXII proteins

Position	Amino acid: Plant	Amino acid: Non-plant	Plant codon	Closest consensus codon(s)	Minimal change to yield consensus codon	Species: Monocots: Maize	Monocots: Rice	Monocots: Wheat	Dicots: Pea	Dicots: Soybean	Dicots: *Oenothera*
42	His	Asp	CAY*	GAY	C→G	+	+	+	+	+	+
87	**Arg**	**Trp**	**CGG**	**UGG**	**C→U**	**+**	**+**	**+**	**+**	**+**	**+**
92	Ser	Ala	AGU	GCU	A→G, G→C	+	+	+	+	+	+
95	**Pro**	**Leu**	**CCN†**	**CUN**	**C→U**	**+**	**+**		**+**	**+**	**+**
111	Gly	Glu	GGG	GAG	G→A	+	+	+	+		
129	**Arg**	**Trp**	**CGG**	**UGG**	**C→U**	**+**	**+**	**+**	**+**	**+**	**+**
156	**Pro**	**Leu**	**CCA**	**CUA**	**C→U**	**+**	**+**	**+**			
162	Gly	Arg	GGU	CGU	G→C				+		
196	**Ser**	**Leu**	**UCA**	**UUA**	**C→U**	**+**	**+**	**+**			**+**
207	**Ser**	**Leu**	**UCA**	**UUA**	**C→U**	**+**	**+**	**+**			
209	Leu	Gln	CUU	CAR	U→A, U→R	+	+	+			
228	**Arg**	**Cys**	**CGU**	**UGU**	**C→U**			**+**			
235	**Thr**	**Met**	**ACG**	**AUG**	**C→U**	**+**	**+**	**+**	‡	**+**	
250	**Arg**	**Trp**	**CGG**	**UGG**	**C→U**				**+**	**+**	**+**

Inferred radical amino-acid substitutions in plant COXII sequences at positions that are invariant in five non-plant species are listed (see Fig. 1). Each amino-acid position is numbered according to the wheat sequence. The plant codons and the closest related codon corresponding to the amino-acid residue at the same position in the non-plant consensus are shown, along with the nucleoside change that would be required at the RNA level to convert the plant codon to the closest consensus codon. Amino-acid substitutions involving a chemical difference (based on polarity, molecular volume and composition[26]) >73 were considered to be nonconservative and were included in the table. +, Site at which a particular plant COXII protein differs from the non-plant consensus (that is, a possible site of RNA editing), with C→U changes indicated in bold. R denotes G or A.

* *Oenothera*, Y denotes U; other plants, Y denotes C.

† Dicots, N denotes G; maize, N denotes A; rice, N denotes U.

‡ Codon not present in the published pea COXII gene sequence[27].

supporting the view that it is involved in copper binding[19]. Our findings predict that examples of selective editing of mitochondrial mRNAs will be found in other plants. For example, conversion of codon 95 (CCA) in the maize COXII coding sequence[12] to either CUA or UUA would result in a proline-to-leucine switch at position 95 in maize COXII, restoring a conserved leucine that occurs at this site in the transmembrane domain of the COXII protein of wheat and other organisms (open triangle, Fig. 1). Finally, our data indicate that the universal genetic code is used in plant mitochondria, with non-edited CGG codons specifying arginine, not tryptophan.

RNA editing phenomena, such as we have described here, challenge the concept of faithful transmission of genetic information from DNA through RNA to protein. For plant mitochondria, it is clear that the gene sequence is an imperfect predictor of the protein sequence: mRNA or direct protein-sequence data, or both, are also required. This raises the important question of where the information introduced during the editing process originates, that is, how the selectivity of the process is achieved. No strong clues about this are provided by the examination of the primary sequence and the potential secondary structure in the vicinity of edited sites in wheat COXII mRNA. Furthermore, the timing of the observed RNA editing relative to that of transcription, its mechanism (for example, deamination versus base deletion–insertion), and its relationship to other forms of RNA editing, particularly the C→U editing of mammalian apolipoprotein-B transcripts[4-7], remain to be examined.

The discovery of yet another instance of RNA editing also raises the issue of just how general this phenomenon is. Protein sequence comparisons indicate that COXII mRNA editing is probably widely distributed among flowering plants (Table 1) and that editing extends to the mRNAs of other genes, such as *coxI* (data not shown). Given that plant mitochondria seem to be of direct eubacterial ancestry[22,23], it is pertinent to ask whether RNA editing may exist among the contemporary eubacterial relatives of plant mitochondria, the closest of which are found within the α-subdivision of the purple bacteria[24,25]. □

Received 31 July; accepted 25 August 1989

1. Benne, R. *et al. Cell* **46,** 819–826 (1986).
2. Simpson, L. & Shaw, J. *Cell* **57,** 355–366 (1989).
3. Benne, R. *Biochim. biophys. Acta* **1007,** 131–139 (1989).
4. Powell, L. M. *et al. Cell* **50,** 831–840 (1987).
5. Chen, S.-H. *et al Science* **238,** 363–366 (1987).
6. Tennyson, G. E., Sabatos, C. A., Higuchi, K., Meglin, N. & Brewer, H. B. Jr *Proc. natn. Acad. Sci. U.S.A.* **86,** 500–504 (1989).
7. Tennyson, G. E., Sabatos, C. A., Eggerman, T. L. & Brewer, H. B. Jr *Nucleic Acids Res.* **17,** 691–698 (1989).
8. Thomas, S. M., Lamb, R. A. & Paterson, R. G. *Cell* **54,** 891–902 (1988).
9. Cattaneo, R., Kaelin, K., Baczko, K. & Billeter, M. A. *Cell* **56,** 759–764 (1989).
10. Shaw, J. M. *et al. Cell* **53,** 401–411 (1988).
11. Feagin, J. E., Abraham, J. M. & Stuart, K. *Cell* **53,** 413–422 (1988).
12. Fox, T. D. & Leaver, C. J. *Cell* **26,** 315–323 (1981).
13. Fox, T. D. *A. Rev. Genet.* **21,** 67–91 (1987).
14. Gualberto, J. M. *et al. Molec. gen. Genet.* **215,** 118–127 (1988).
15. Bland, M. M., Levings, C. S. III & Matzinger, D. F. *Molec. gen. Genet.* **204,** 8–16 (1986).
16. Swamy, G. S. & Pillay, D. T. N. *Plant Sci. Lett.* **25,** 73–84 (1982).
17. Maréchal, L. *et al. Nucleic Acids Res.* **13,** 4411–4416 (1985).
18. Bonen, L., Boer, P. H. & Gray, M. W. *EMBO J.* **3,** 2531–2536 (1984).
19. Brunori, M., Antonini, G., Malatesta, F., Sarti, P. & Wilson, M. T. *Adv. inorg. Biochem.* **7,** 93–153 (1988).
20. Geliebter, J. *et al. Proc. natn. Acad. Sci. U.S.A.* **83,** 3371–3375 (1986).
21. Steffens, G. J. & Buse, G. *Hoppe-Seyler's Z. physiol. Chem.* **360,** 613–619 (1979).
22. Gray, M. W. & Doolittle, W. F. *Microbiol. Rev.* **46,** 1–42 (1982).
23. Gray, M. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **86,** 2267–2271 (1989).
24. Yang, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **82,** 4443–4447 (1985).
25. Cedergren, R. *et al. J. molec. Evol.* **28,** 98–112 (1988).
26. Grantham, R. *Science* **185,** 862–864 (1974).
27. Moon, E., Kao, T-h. & Wu, R. *Nucleic Acids Res.* **13,** 3195–3212 (1985).
28. Grabau, E. A. *Curr. Genet.* **11,** 287–293 (1987).
29. Feng, D.-F. & Doolittle, R. F. *J. molec. Evol.* **25,** 351–360 (1987).
30. Auchincloss, A. H. & Brown, G. G. *Biochem. Cell Biol.* **67,** 315–319 (1989).
31. Bonen, L. & Gray, M. W. *Nucleic Acids Res.* **8,** 319–335 (1980).
32. Schnare, M. N., Spencer, D. F. & Gray, M. W. *Can. J. Biochem. Cell Biol.* **61,** 38–45 (1983).
33. Macdonell, M. T., Hansen, J. N. & Ortiz-Conde, B. A. *Meth. Microbiol.* **19,** 357–404 (1987).
34. Tabor, S. & Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **84,** 4767–4771 (1987).
35. Kao, T-h., Moon, E. & Wu, R. *Nucleic Acids Res.* **12,** 7305–7315 (1984).

ACKNOWLEDGEMENTS. We thank L. Bonen (University of Ottawa) for sharing DNA clones and sequence information, D. F. Spencer, M. B. Coulthart, P. J. Hanic-Joyce, M. J. Covello and K. Conover for comment and assistance, and Lisa Laskey for help in manuscript preparation. We gratefully acknowledge support from the Medical Research Council of Canada (P.S.C. and M.W.G.). M.W.G. is a Fellow in the Evolutionary Biology Program of the Canadian Institute for Advanced Research.

Kluyveromyces lactis toxin does not inhibit yeast adenylyl cyclase

John H. White, Andrew R. Butler & Michael J. R. Stark

Department of Biochemistry, The University, Dundee DD1 4HN, UK

KILLER strains of the yeast *Kluyveromyces lactis* secrete a complex protein toxin, production of which is associated with the presence of two linear double-stranded DNA plasmids, pGKL1 and pGKL2[1]. The toxin is encoded by two of the four genes carried by pGKL1[2] and is active against yeast from several genera (including *Saccharomyces cerevisiae*)[1,3]. The toxin arrests sensitive yeast cells in the unbudded (G_1) phase of the cell cycle[4], indicating that it could be used to study stage-specific events occurring in G_1. Previous work indicated that *K. lactis* toxin was a reversible inhibitor of adenylyl cyclase in sensitive *S. cerevisiae* cells[4]. Here we report that although the toxin does induce G_1 arrest, its effects on *S. cerevisiae* cells cannot be mediated by inhibition of adenylyl cyclase, and that treated cells never recover from its effects.

The addition of toxin to a culture of sensitive *S. cerevisiae* cells leads to a cessation of cell division within 1–2 generations, as indicated by the complete block in accumulation of cells after treatment with toxin (Fig. 1*a*). In treated cultures, we saw no further increase in cell numbers for over 35 h. Concomitant with this inhibition, we found that the proportion of budded cells in the population fell rapidly from ≳45% to ~10% (Fig. 1*c*). Because G_1 is the only phase of the cell cycle during which cells are unbudded, this clearly indicated that cells become arrested in G_1 after the addition of toxin. Treated cells, however, continue to increase in volume without substantial inhibition of protein and RNA synthesis (A.R.B., J.H.W. and M.J.R.S., manuscript in preparation), indicating that some degree of growth is possible in the arrested state. Previous work[4] indicated that the toxin had no effect on cell viability. But we found that the toxin caused a rapid and progressive loss of viability over the first 20 h, which was sufficient to explain the block to cell division (Fig. 1*b*). We used a range of toxin concentrations (Fig. 2) and were unable to find any conditions under which the arrest in cell division could be obtained without the loss of viability; concentrations of toxin too low to affect viability simply failed to arrest cell division. Although the treated population eventually seemed to regain viability (Fig. 1*b*), this was not due to the recovery of the inhibited cells, but rather resulted from the growth of spontaneously arising toxin-resistant cells, which are present in the starting population at a frequency of $\sim 2\times10^{-6}$. The regain of viability thus preceded the resumed increase in total cell numbers, and the change in cell numbers predicted from the viability data closely mirrors that observed experimentally (Fig. 1*a*). By the end of the experiment, all viable cells in the treated population are resistant to toxin (data not shown). In addition, treatment with toxin leads to a sixfold enrichment for respiratory deficient cells, although respiratory deficiency *per se* does not cause toxin resistance (data not shown). Moreover, the non-viable cells in the treated population were able to maintain a membrane potential, indicated by their failure to accumulate vital stains such as methylene blue or Janus green. This is in complete contrast to other yeast killer toxins, which act as ionophores[5]. Because the toxin is therefore an irreversible inhibitor, it is impossible to perform reciprocal shift experiments to map its execution point in the cell cycle. When cells are arrested in S phase with hydroxyurea, treated with toxin, and then the hydroxyurea block released in the continued presence of toxin, most of the budded cells are able to complete a round of division and become arrested in the unbudded state (data not shown). This indicates that the toxin-sensitive step occurs before the S phase. We therefore conclude that toxin leads to a permanent arrest of cells in G_1, such that they can never resume mitotic division.

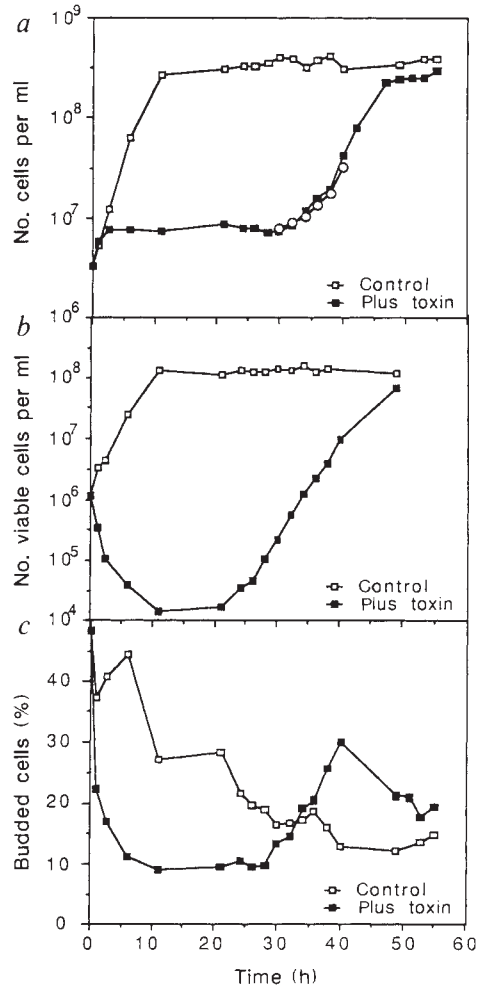


FIG. 1 Effect of the *K. lactis* toxin on cell division, cell viability and budding index. Cells of LL20 (*MATα leu2-3, leu2-112, his3-11, his3-15*) were grown overnight in YPD (per l:10g yeast extract, 20g peptone, 20g dextrose) medium[10], subcultured at 10^6 cells per ml into fresh medium and shaken in flasks at 28 °C. When the cell density reached 3×10^6 per ml (time zero), the culture was split into two 20 ml portions using pre-warmed flasks, one of which received 5 mg l^{-1} purified *K. lactis* toxin[2] (~30 μM). This concentration of toxin was the minimum that gave complete inhibition of growth as measured using a microtitre well assay[11]. Samples from each culture were examined at regular intervals for total cell number per ml, number of viable cells per ml and the proportion of budded cells. Viable cells were determined by plating suitable dilutions of cells (in YPD) on YPD agar and incubating at 28 °C. □, Control; ■, plus toxin; ○, predicted curve for apparent recovery (see text).

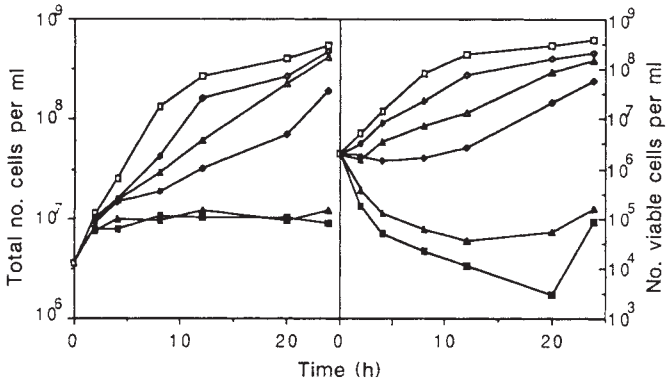


FIG. 2 Dose-response of sensitive cells to toxin. The experimental conditions were as described in Fig. 1 legend, except that several different toxin concentrations were used. □, No toxin; ◇, 250 ng l^{-1} toxin; △, 500 ng l^{-1} toxin; ◆, 750 ng l^{-1} toxin; ▲, 5 mg l^{-1} toxin; ■, 10 mg l^{-1} toxin.

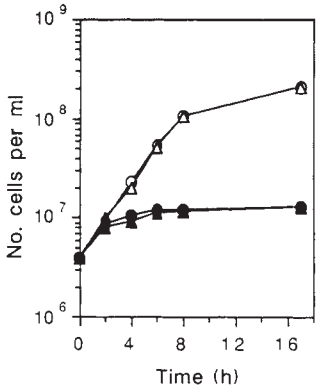


FIG. 3 Failure of cAMP and IBMX to antagonize toxin action. The experimental conditions were as described in Fig. 1 legend, except that strain JR242-17D (*MATα his3 ura3 trp*$^-$) was used and some cultures received cAMP and IBMX at time zero. ○, Control; ● and ▲, plus toxin (5 mg l^{-1}); △ and ▲, plus cAMP (5 mM) and IBMX (5 mM).

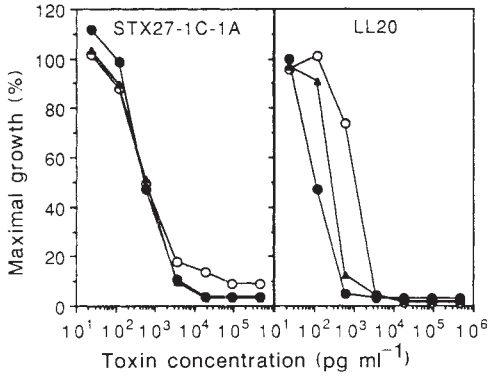


FIG. 4 Disruption of the *BCY1* gene does not render cells toxin-resistant. Growth of strains LL20 and STX27-1C-1A[4] was measured using a microtitre well assay[11] and growth of 10^5 cells in the presence of differing toxin concentrations expressed as a percentage of growth without toxin. The *bcy1::LEU2* disruptants derived from each strain were generated by standard methods using a previously described construct[8] and checked by Southern-blot analysis to confirm that the wild-type *BCY1* locus had been replaced. The *bcy1::LEU2* strains were further checked for heat-shock sensitivity, sensitivity to nitrogen starvation and failure to grow on acetate as a carbon source, all of which are characteristic of the *bcy1* mutant phenotype. ○, Parent *BCY1*$^+$ strains; ● and ▲, *bcy1::LEU2* disruptants derived from the parent strains.

We then re-examined the hypothesis that *K. lactis* toxin acts by inhibiting yeast adenylyl cyclase[4]. Previous work using strains STX27-1C-1A and F10D indicated that exogenous cyclic AMP antagonized inhibition by the toxin[4]. By using JR242-17D (Fig. 3), STX27-1C-1A or F10D, however, we have been unable to demonstrate any effect of added cAMP, alone or together with the phosphodiesterase inhibitor 3-isobutyl-1-methylxanthine (IBMX), on the efficacy of the toxin (Fig. 3, and data not shown). Moreover, we have used a genetic approach to demonstrate that the toxin does not act by preventing synthesis of cAMP. Because the only essential roles of cAMP in mitotic cell growth and division are mediated by cAMP-dependent protein phosphorylation, cells in which the catalytic subunits of the cAMP-dependent protein kinases are constitutively active no longer require cAMP to undergo growth and division[6,7]. This condition is readily produced by genetic disruption of the *BCY1* gene, the single locus that encodes the regulatory subunit of the cAMP-dependent protein kinases[7]. In fact, *bcy1* mutants[8] or strains with a disruption in the *BCY1* gene (J. Broach, personal communication) can grow in the absence of a functional gene encoding adenylyl cyclase. We have therefore generated derivatives from several yeast strains (including LL20 and STX27-1C-1A) in which the wild-type *BCY1* gene had been disruped by replacement with a *bcy1*::*LEU2* gene construct, and tested these for their response to toxin. In every case, these derivatives were fully sensitive to the toxin and showed growth arrest and loss of viability just as observed with the parent stains (Fig. 4, and data not shown). This indicates that toxin cannot exert its effects by inhibition of adenylyl cyclase. The mode of action of the toxin is therefore still unknown and it may yet prove a useful tool for examining stage-specific events in G_1. We anticipate that the toxin-resistant mutants that we and others[9] have isolated will provide the most direct route to characterizing the target of the toxin. □

Received 20 July; accepted 8 September 1989.

1. Gunge, N., Tamaru, A., Ozawa, F. & Sakaguchi, K. *J. Bact.* **145**, 382–390 (1981).
2. Stark, M. J. R. & Boyd, A. *EMBO J.* **5**, 1995–2002 (1986).
3. Gunge, N. & Sakaguchi, K. *J. Bact.* **147**, 155–160 (1981).
4. Sugisaki, Y., Gunge, N. Sakaguchi, K., Yamasaki, M. & Tamura, G. *Nature* **304**, 464–466 (1983).
5. Kagan, B. L. *Nature* **302**, 709–711 (1983).
6. Matsumoto, K., Uno, I., Oshima, Y. & Ishikawa, T. (1982). *Proc. natn. Acad. Sci. U.S.A.* **79**, 2355–2359 (1982).
7. Toda, T. *et al. Molec. cell. Biol.* **7**, 1371–1377 (1987).
8. Kataoka, T., Broek, D. & Wigler, M. *Cell* **43**, 493–505 (1985).
9. Tokunaga, M. Kawamura, A. & Hishinuma, F. *Nucleic Acids Res.* **17**, 3435–3446 (1989).
10. Sherman, F., Fink, G. R. & Hicks, J. B. *Methods in Yeast Genetics* (Cold Spring Harbor, New York, 1986).
11. Romanos, M. A. & Boyd, A. *Nucleic Acids Res.* **16**, 7333–7350 (1988).

ACKNOWLEDGEMENTS. We thank Prof. James Broach for providing strains JR242-17D, STX27-1C-1A and F10D and Dr Alan Wheals for discussions. The *bcy1*::*LEU2* construct was produced in the laboratory of Prof. M. Wigler. This work was supported by a project grant from the MRC (M.J.R.S.). J.H.W. is the recipient of a research studentship from SERC.

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse. I am interested in electronic ordering through the following system(s):

☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnTyme ☐ OCLC ILL Subsystem
☐ Other (please specify)_____
☐ I am interested in sending my order by mail.
☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name_____
Title_____
Institution/Company_____
Department_____
Address_____
City_____ State____ Zip_____
Phone (____)_____

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

Free-flow electrophoresis

H. Wagner

The continuous and simultaneous separation and fractionation of samples by free-flow electrophoresis has both analytical and preparative applications.

PIONEERING work by Hannig[1] led to the development of free-flow electrophoresis, also known as continuous-flow electrophoresis, as a technique for the purification of biopolymers, which now has broad applications in biology and medicine.

During free-flow electrophoresis, a thin film of electrolyte solution flows laminarly between two parallel cooling plates and an electric field is applied perpendicular to the flow, resulting in the differential deflection of charged solutes and particles. Depending on the mode of separation used, these substances are separated according to differences in either electrophoretic mobility or isoelectric point. The input of electrolyte solution and sample at one end of the electrophoresis chamber and fractionation at the other end are carried out continuously.

While free-flow electrophoresis can be used for analytical purposes, its continuous character lends itself to preparative applications. The need to solve purification problems in biotechnology has spurred interest in the preparative aspects of this technique[2].

Modes of separation

Free-flow electrophoresis has several separation modes[3] (Fig. 1), including zone electrophoresis[4], isotachophoresis[5], isoelectric focusing[6] and field-step electrophoresis[7].

Zone electrophoresis was the first free-flow electrophoresis separation mode to be developed. In this mode, a narrow sample stream is injected into a uniform background electrolyte, which usually has a constant pH and conductivity. Charged sample compounds are deflected by the electric field according to their electrophoretic mobilities. However, band broadening and dilution of the separated compounds can occur due to the laminar flow profile, diffusion and electro-osmosis. In spite of this, zone electrophoresis is still used for many applications, particularly in the separation of living cells.

The isotachophoresis mode is used for the separation of either positively or negatively charged particles. The sample is introduced into the separation chamber as a parallel zone between a leading and a terminating electrolyte. The mobilities of the sample compounds must be intermediate between that of the leading ion, which has the highest mobility, and the terminating ion, which has the lowest. The counter ions are chosen to have a high buffering capacity. If the steady state is achieved, all sample components move with equal velocity and are separated into individual zones. Ideally, there are as many zones as sample compounds. These adjoin without any spaces and in order of decreasing mobility. The concentration of the separated compounds depends only on the concentration of the leading ion and may be higher in the zone than in the original mixture.

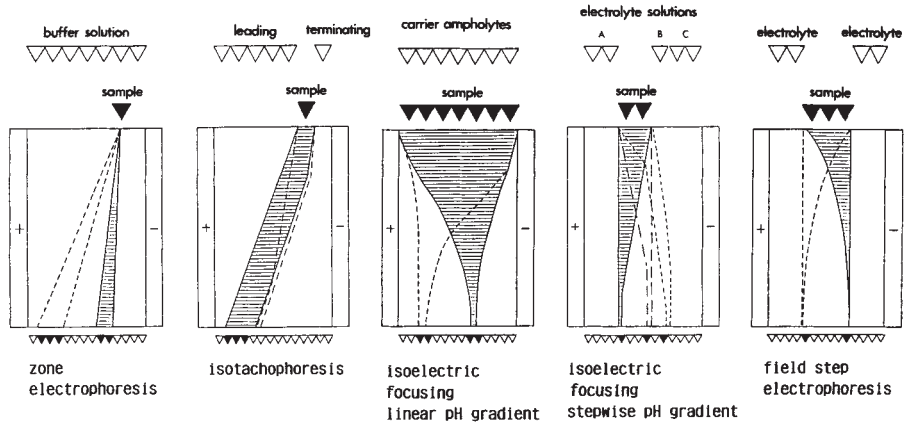


FIG. 1 Schematic representation of the different modes of separation with free-flow electrophoresis. ▽, background buffer; ▼, sample.

In isoelectric focusing mode, low molecular weight amphoteric substances — the carrier ampholytes — are introduced into the chamber, forming a linear pH gradient and causing proteins to be separated according to their isoelectric points. The pH gradient can be generated electrophoretically within the system by introducing a complex mixture of ampholytes. The sample can then be introduced either uniformly mixed with the ampholytes or as a narrow zone within the ampholyte mixture. A slow flow rate is necessary as the pH gradient must form and the sample compounds focus during the residence time.

As an alternative to forming the gradient electrophoretically, a stepwise, preformed pH gradient may be admitted. This avoids its time-consuming generation by the electric field. The stepwise pH gradient can be produced using a few solutions of pure amphoteric substances. The number of pH steps is given by the number of different solutions introduced. The sample is mixed and admitted with one of the solutions. This method allows the isolation of a single substance from a complex mixture.

In field-step electrophoresis mode, separation takes place in a discontinuous electrolyte system in which the sample solution with low ionic strength is flanked by two electrolyte solutions of high ionic strength. The conductivities of the two flanking electrolyte solutions must be about 20 times higher than that of the sample solution[7]. This results in a step-like profile for the field strength. Charged substances within the sample will migrate towards the respective electrodes according to electrical charge, but at the boundary with the solution of high conductivity, the migration velocity will be greatly reduced. Thus, focusing at the conductivity boundaries is due to the change in migration velocity, and not the loss of electrical charge, as with isoelectric focusing.

In spite of this, the direction of migration can be regulated by choosing a suitable pH when separating amphoteric substances such as proteins or peptides. However, as enrichment factors of only 10–15 are achieved during focusing, field-step electrophoresis is used mainly for the preseparation of complex mixtures.

Focusing using conductivity boundaries can be used in combination with isoelectric focusing. Although such methods need to be optimized for the specific separation problem, they result in high throughput and concentration of the sample.

Documented applications

The various applications of free-flow electrophoresis are documented in more than 500 publications, most of which are cited in review articles[3,4,8-10]. This technique has been used to separate substances of vary-

ing size, from small ions to whole living cells. With the latter, low electrostatic forces and the absence of carrier materials are a particular advantage. Successful separations on a laboratory scale of cells, subcellular particles, viruses and bacteria have been documented. Recently, free-flow electrophoresis has been used for the separation of peptides and proteins, including the separation and purification of single components from a complex biological mixture[2,11,12].

The pH, conductivity and composition of the electrolyte solutions can be freely selected within reasonable limits. The continuous nature of free-flow electrophoresis as opposed to most other physicochemical separation methods makes it suitable for not only analytical but also preparative separations, although scale-up is seriously hindered by the unavoidable generation of Joule heat. Despite this, advancements in the computerization of process control makes free-flow electrophoresis an attractive option for the pharmaceutical industry.

Instrumentation

In addition to the separation chamber, other components required for a complete free-flow electrophoresis instrument include an efficient cooling device with precise temperature control, a high-voltage power supply with a maximum capacity of 1,000 W, and high-precision pumps for the input of electrolyte and sample and for fractionation. An electronic unit controls the entire process and all peripheral equipment in conjunction with an online detection system for analysis of the separation profile during the run.

Various free-flow electrophoresis instruments have been developed but few are commercially available. One such instrument is the Elphor VaP 22 developed by Bender & Hobein, Munich, FRG, but instruments are also available from Hirschmann, Munich, FRG, and Protein Technologies, Tucson, Arizona. □

Professor Horst Wagner is at the Department of Inorganic Analytical Chemistry & Radiochemistry, University of Saarland, Em Stadtwald, 6600 Saarbrucken, FRG. For more information, fill in reader service number 100.

1. Hannig, K. *Analyt. Chem.* **181,** 244 (1961).
2. Kuhn, R. & Wagner, H. *Electrophoresis* **10,** 165 (1989).
3. Wagner, H., Kuhn, R. & Hoffstetter, S. *Praxis der elektrophoretischen Trennmethoden.* (Springer, Heidelberg 1989).
4. Hannig, K. *Electrophoresis* **3,** 235 (1982).
5. Schmitz, G., Bottcher, A., Kahl, H.G. & Bruning, T. *J. Chromatog.* **431,** 327 (1988).
6. Wagner, H. & Speer, W. *J. Chromatog.* **157,** 366 (1978).
7. Wagner, H. & Kessler, R. *GIT Lab. Med.* **7,** 30 (1984).
8. Hannig, K. & Heidrich, H.G. *Free flow electrophoresis: An important analytical and preparative method of biology, biochemistry and diagnostics.* (GIT Verlag, Ernst Giebeler, Darmstadt 1989).
9. Morre, D.J., Howell, K.E. & Cook, G.M.W. *Cell free analysis of membrane traffic.* (A.R. Liss, Inc., New York 1988).
10. Marsh, M. *et al. J. Cell Biol.* **104,** 875 (1987).
11. Kuhn, R., Wagner, H., Mosher, R.A. & Thormann, W. *Electrophoresis* **8,** 503 (1987).
12. Kuhn, R. & Wagner, H. *J. Chromatog.* **481,** 343 (1989).

Molecules on the move

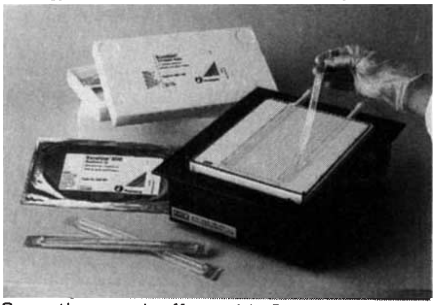
Microwave-in-the-bag agarose, a scanning densitometer with zig-zag optics, and an automated DNA sequencing system are among this week's products that focus on electrophoresis.

AMERSHAM Corp. is now offering **pre-labelled [^{35}S] DNA molecular weight markers** for use as standards in hybridization techniques, such as Southern blotting *(Reader Service No. 101)*. Each set of markers contains 17 fragments that range from 0.6 to 22.01 kb. The pre-labelled markers are supplied ready-to-use and, says Amersham, offer high resolution and an even distribution of fragments.

GELSIM is Biosoft's agarose gel **electrophoresis simulation and analysis package** which allows the user to enter banding patterns obtained from a set of running conditions and determine how the patterns change when the conditions are altered (*Reader Service No. 102*). The adjustable running conditions include voltage, ionic strength of the running buffer, agarose gel concentration, duration of run, electrode separation and agarose gel length. The 'electronic' agarose gel provides eight lanes, each of which can hold up to 30 fragments. DNA fragments can be entered by size via the keyboard, or by position on the gel by moving the cursor. Features of the package include 30 pop-up help screens, banding pattern analysis, DNA size estimation, facility for printed gel summary sheets and hard-copies of the gel and banding patterns, and a diskette library of over 270 restriction digest fragment sets. The £99 (UK) GELSIM software is available on a 5.25-or 3.25-inch disc for use on an IBM PC.

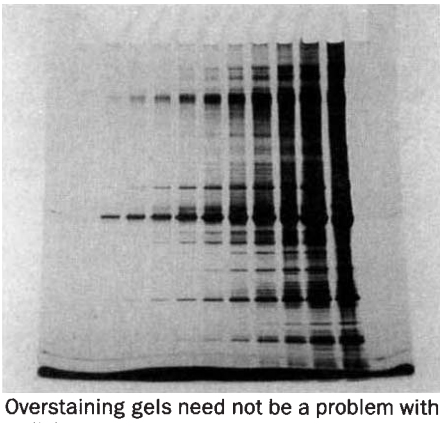
Ready-to-use reagents

As an alternative to casting gels and mixing buffers, and for the rapid separation of SDS-denatured proteins in the molecular weight range of 6,500 to 300,000 daltons, Pharmacia LKB recommends its **precast SDS gradient gels and buffer strips** (*Reader Service No. 103*). Designed for use with the Multiphor and Ultrophor IEF systems as well as other horizontal systems, the precast ExcelGel SDS, Gradient 8-18 gel media and precast ExcelGel SDS Buffer Strips eliminate the need for the preparation of toxic gel solutions, time-consuming gel casting and buffer preparation, says Pharmacia. Up to 52 samples or two 2-D maps can be run on each 110 × 245-mm gel, says the company. ExcelGel SDS, Gradient 8-18 gel media is supplied in packages of six plastic-backed gels and ExcelGel SDS Buffer Strips in packages of six anode and six cathode strips.



Save time and effort with Pharmacia LKB's precast SDS gradient gels and buffer strips.

According to Daiichi Pure Chemicals Co. Ltd, overstaining gels could be a thing of the past with their new 2-D Silver Stain II **silver staining kits** (*Reader Service No. 104*). The kits can be used to stain plastic-backed gels, IEF gels as well as regular gels, providing, says Daiichi, uniform staining of proteins with a high signal-to-noise ratio. The company says the procedure could not be easier: the 'stop' solution is added after 10 minutes of reduction and so there is no need to stand over the gel during the developing process. Each kit comes complete with six 100-ml plastic bottles containing sufficient reagents to stain 10 regular-sized gels or 20 minigels, says Daiichi.



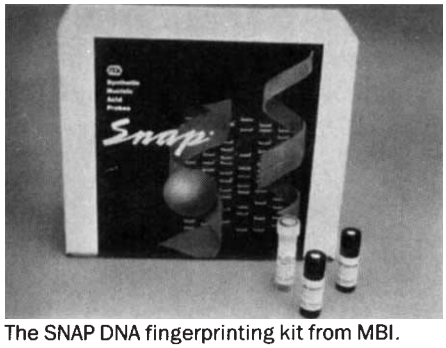
Overstaining gels need not be a problem with Daiichi's silver staining kit.

Cut down on reagent preparation with the convenience of **DNA agarose in microwavable pouches** from Integrated Separation Systems (*Reader Service No. 105*). Preparation is simple — just snip the corner of the pouch and microwave for 1-2 minutes. Each pouch is sufficient for two 3-cm thick minigels, says ISS. The company offers a custom-blending service, where they will prepare gels with 0.5 to 1.5 per cent agarose with either Tris-

Borate-EDTA or Tris-Acetate-EDTA buffers. ISS says all reagents are tested for purity and reproducibility of results.

DNA fingerprinting

Non-radioactive DNA fingerprinting technology is now available in kit form with the SNAP DNA Fingerprinting Kit from Molecular Biosystems, Inc. (*Reader Service No. 106*). The SNAP kit has applications in tissue typing, identification and



The SNAP DNA fingerprinting kit from MBI.

paternity studies, gene mapping and quality control of inbred or transgenic animals, says MBI. Hybridization to the complementary sequence in the target nucleic acid is achieved using synthetic oligonucleotide probes that are 20 to 30 bases long. Direct conjugation of alkaline phosphatase to the probe simplifies the protocol and reduces non-specific binding of the enzyme, which can produce increased background, says the company. Detection of hybridized target nucleic acid is by a one-step colour development process that results in localized blue staining. MBI says the assay takes less than four hours and each $195 (US) kit contains 50 pmols of probe — enough for about 300 hybridizations.

DNA fingerprinting probes for research use are now available from Cambridge Research Biochemicals (*Reader Service No. 107*). Developed in collaboration with Cellmark Diagnostics, the stated applications include verification of tissue culture cell-line purity, human chromosome mapping, plant and animal breeding and population genetics studies. The company is making available multi-locus DNA probes (33.15 and 33.6) and single-locus probes (MS1, MS31, MS43, MS8, and g3). The probes are full-length, purified DNA inserts, and are supplied as 60-ng aliquots suitable for direct labelling. Each vial comes with a Technical Data Sheet and sufficient material for the analysis of up to 50 genomic DNA samples on up to 10 membranes, says CRB.

Sequencing systems

For the quick and easy **preparation of nucleic acid sequencing gels**, National Diagnostics recommends its ready-to-use SEQUAGEL Sequencing System (*Reader Service No. 108*). Using Sequagel, which consists of acrylamide, bisacrylamide, urea and TBE buffer, gels from 0–20 per cent monomer can be prepared. The time-consuming and potentially hazardous process of weighing, mixing and filtering reagents is eliminated, says the company. The Sequagel solution is stabilized by the incorporation of a gaseous inhibitor, which, National Diagnostics says, prevents the 'self-polymerization' to which solid acrylamide is susceptible. The gaseous inhibitor is removed upon deaeration of the gel casting solution. Oxidation products such as aldehydes, which can alter migration, are removed by selective adsorption, producing consistent R_F values from run to run, says the company.

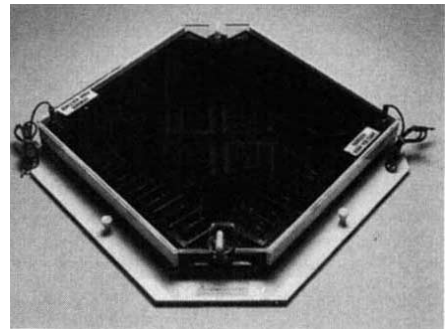
A re-engineered gel electrophoresis system and upgraded sequencing software are new features of the Acugen 402 **automated DNA sequencer** from EG & G Biomolecular (*Reader Service No. 109*). The Acugen 402 is, says the company, competitively priced at $35,000 (US), and uses standard radioisotopic labelling techniques (^{32}P) for the direct determination of DNA sequences at or above 99 per cent accuracy. The system's redesigned gel cassette improves the gel casting process and eliminates gel aberrations such as

Automated DNA sequencing with the upgraded Acugen 402 from EG & G Biomolecular.

lane drift, gel 'smiles' and 'frowns', says EG & G. The system can sequence 3,600 bases per day in two runs, says the company, and features the online viewing of raw data via a high-resolution monitor. The PC-compatible GENQUEST sequence analysis software has been upgraded to include more functions for base sequence determination and analysis.

Hardware options

Pulsed-field mapping and the separation of multi-megabase DNA fragments are possible with the OSP-20CP **constrained uniform field electrophoresis system** (CUFE) from Owl Scientific Plastics, Inc. (*Reader Service No. 110*). The CUFE system functions independently of voltage, gel thickness and buffer conductivity and can, says the company, be run at high voltages for rapid separations. No external circuitry is involved as 'baffles' physically constrain the electric field generated from platinum electrodes. The $3,695 (US) OSP-20CP comes complete with



Owl Scientific Plastics' constrained uniform field electrophoresis system.

ADVERTISEMENTS

for IBM-compatible personal computers

Software for Your Lab

CELLINE Keeps accurate records of cell lines, cultures, freezes and ampule storage locations in all LN_2 and mechanical freezers.

GELMETRIC Calculates DNA and protein sizes from gel mobilities. Use with or without a digitizer.

SMGWS 2 Draws stereoscopic 3-D space-filling macromolecules on a PS/2 computer!

Money-Back Guarantee!

BioScience SOFTWARE (814)692-7204

P.O. Box 10003, State College, PA 16805

Reader Service No.26

CUSTOM-MADE & PURIFIED
DNA PEPTIDES

$8/base
50–100 µg
40 hour delivery

Quotes available
50–300 mg
1–3 wks delivery

GUARANTEED QUALITY & QUANTITY

V.G. CORP.
1200 Wonderland Rd. S.
Bldg #7, Unit #2
London, Ontario
Canada N6L 1A8

CALL OR FAX: (519) 652-3758

Reader Service No.29

Custom DNA

Purified and Delivered in 48 hours

100 ug, $10.00 a base for the first 15 bases, $7.50 for each additional base. Milligram quantities available.

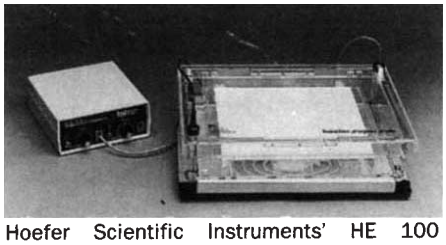
Research Genetics
2130 S. Memorial Parkway Huntsville, AL 35801

1-800-533-4363
In the United Kingdom 0-800-89-1393
In Canada 1-800-833-8348
FAX 205-536-9016 (NW2971)D

Reader Service No.15

programmable power inverter, recirculation system, gel casting tray that can accommodate 20 × 20-cm gels, comb and protocol.

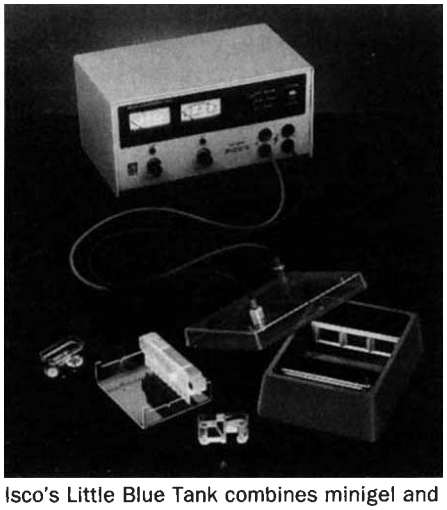
The safe internal recirculation of buffer and a built-in cooling system are features of the HE 100 SuperSub — Hoefer Scientific Instruments' **horizontal electrophoresis unit** *(Reader Service No. 111)*.



Hoefer Scientific Instruments' HE 100 SuperSub with buffer recirculation and built-in cooling system.

Buffer is continuously recirculated through a spiral coil built into the gel platform. When placed on a magnetic stirrer, a magnetic spinbar in the recirculation coil causes the buffer to flow over the top of the gel, which, says Hoefer, avoids ion depletion, pH changes and provides a uniform temperature. The design also features a serpentine heat exchange coil through which coolant can be pumped, simultaneously cooling the underside of the gel and the buffer flowing over the gel, says the company. The HE 100 can be fitted with a wide range of running plates for gels that are 20-cm wide and 15-, 20-, and 25-cm long.

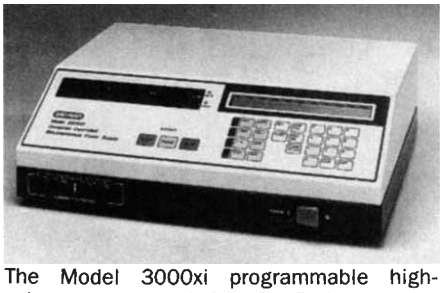
The Little Blue Tank from Isco, Inc. is a multifunctional electrophoresis system that **combines minigel and electroelution techniques** into one system *(Reader Service No. 112)*. The $295 (US) system, not including the power supply, comprises a horizontal buffer tank, a 9 × 10-cm UV-transparent gel tray, combs for 12-lane analytical and 3-lane preparative gels, and two types of electroelution accessories. The electrode design provides, says Isco, a uniform electric field and run times that take 30–60 minutes at 50 V. Replacing the gel tray with up to four membrane-trap electroelution cups and varying the molecular weight cut-off point of the membrane between 1,000 and 50,000, provides a means of eluting proteins and other macromolecules from gels and solutions. Elution times for proteins are typically 1–4 hours with 80–100 per cent recovery, says Isco. Nanogram amounts of DNA can be recovered from gel bands by using the Nanotrap electroelution cups, says the company. The Little Blue Tank can also be used for concentrating biomacromolecules from solutions.



Isco's Little Blue Tank combines minigel and electroelution techniques into one system.

Power supplies

The Model 3000xi is the new programmable, high-voltage **electrophoresis power supply** from Bio-Rad Laboratories *(Reader Service No. 113)*. The unit can provide constant voltage to 3,000 V, current to 300 mA and power to 400 W. The 3000xi's 400-W capacity allows the unit to deliver 3,000 V and 150 mA simultaneously, says Bio-Rad. The unit has four operating modes: the three-level step function for programming up to three sets of running conditions; the time-controlled mode for setting parameters for automatic operation; the volt-hour controlled mode which terminates a run after a preset number of volt-hours; and the memory operation mode for storing up to 10 sets of running conditions. At the touch of a button, the power conditions can be recalled from memory, ensuring, says Bio-Rad, reproducible operation in their electrophoresis cells and the precise control of runs.

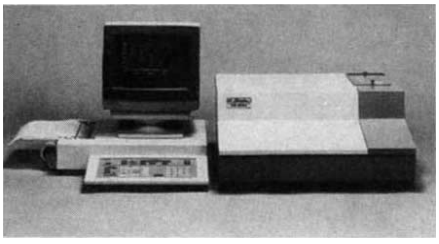


The Model 3000xi programmable high-voltage power supply from Bio-Rad.

For separating kilobase and megabase DNA in pulsed-field or field-inversion chambers, the E-C Apparatus Corp. recommends its EC720 **alternating controlled electrophoresis unit** *(Reader Service No. 114)*. The EC720 microprocessor-controlled unit can connect to any standard electrophoresis power supply and can run up to four chambers simultaneously. The unit supplies a voltage up to 1,500 V and current up to 500 mA. Ramped, stepped or compound changes can be programmed into the run. In addition, the unit will automatically compute the average forward voltage, watts and milliamps. The EC720 features an LED display of throughput voltage, and a 16-character four-line LCD display with a 20-pad input keyboard.

Detection and analysis

According to V.A. Howe & Co. Ltd, the latest optics and computer technology have been combined to produce the Shimadzu CS-9000 dual wavelength **scanning densitometer** for gels, TLC plates and autoradiograms *(Reader Service No. 115)*. The £15,490 (UK) CS-9000 features 'zig-zag' scanning optics, which, the company says, eliminates errors caused by the irregular shape of a chromatographic spot, without significantly affecting scan speed. The instrument can function in transmission, reflection and fluorescence modes, and has a measuring wavelength range of 200–650 nm and a resolution of up to 50 microns, says Howe. Other features include a trace scan function for automatically tracking non-linear separations, background correction function, and colour video display with menu-driven multipage system. Instrument parameters, automatic programs and calibration curves can be stored in the 16-Kbyte memory and hard copies can be printed via the parallel head printer.



The Shimadzu CS-9000 'flying spot' scanning densitometer from V.A. Howe.

Southern-Light Plus from Tropix is the new **chemiluminescent detection and labelling system** for analysing membrane-bound DNA from Southern blots and other DNA hybridizations *(Reader Service No. 116)*. The test kit features the chemiluminescent one-step substrate AMPPD, which is directly activated by alkaline phosphatase, and which can detect as low as 10^{-20} moles of enzyme, says Tropix. The Southern-Light visualization reagents can also be used with biotinylated and digoxigenin-labelled probes. Results are imaged on X ray or instant film within a 5–60 minute incubation period at room temperature, says the company. Each kit consists of Southern-Light visualization reagents and the Trans-Light system for biotinylating DNA using nick translation, and contains sufficient reagents for 50 10 × 10-cm nylon blots. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.